

Making Competency-Based Learning Visible

What Learners Experience Across Ten High Schools

Dr. Rae Lymer, Dr. Jilliam Joe, and Kelly Organ

Recommended Citation

Lymer, R., Joe, J., Organ, K., and Gitiria, L. (2026). Making Competency-Based Learning Visible: What Learners Experience Across Ten High Schools. FullScale.

About FullScale

FullScale's mission is to unite education leaders and organizations together to drive collective learning, action, and systems transformation. We break silos, bridge divides, lift up new evidence and unseen innovations, and accelerate change for lasting impact for every learner.

Table of Contents

Recommended Citation	2
About FullScale	2
Table of Contents	2
Competencies Expand How Learning Looks	4
Different CBL Models Reveal Shared Learner Experiences	5
No Single Pathway to Adopting CBL	7
Applying a Common Lens to Understand Different Models	9
Three Common Learner Experiences Across Ten Different Models	12
The Full Set of Markers Strengthens How Competencies Shape the Learner Experience	18
The Role Competency-Based Learning Plays Within Other Learner-Centered Approaches	19
High Schools Demonstrate Strong Performance on Conventional Success Measures	21
Local Competency Evidence Relates to Conventional Academic Outcomes Measures	23
Competency Evidence Must Travel Across Systems	25
Shared Language Creates a Foundation for a Stronger CBL Evidence Base	26
What's Coming Next	27
References	28
Appendix A. Study Design, Sample, and Data Collection	47
Appendix B. CBL Marker Development, Definitions, and Hypothesized CBL Continuum	49
Appendix C. Psychometric Analysis for Observation Protocol	58

Appendix D. Psychometric Analysis of the CBL Student Survey	73
Appendix E. Psychometric Analysis of the CBL Teacher Survey	92
Appendix F. Learner Focus Group Methods and Analysis	111
Appendix G. Policy Focus Groups, Interviews, and Documentary Analysis	118
Appendix H. Competency and Academic Outcome Methods	123
Appendix I. Student Survey Instrument	132
Appendix J. Teacher Survey Instrument	140

Across the country, high schools are leveraging competency-based approaches to transform the student experience, but little consensus exists as to what that looks like in practice.

Underlying these efforts is a powerful premise: learners are more likely to reach their full potential when learning is personalized to who they are,¹ competency-based in how they progress,² and whole-child in what they develop.³ The challenge is translating that vision into learning experiences that are tangible for young people and workable within the realities of high school. To better understand what that looks like for the learners at the center of these efforts, FullScale conducted a study during spring 2026 to make competency-based learning (CBL) experiences more visible.

A [national landscape scan](#) of high schools applying various student-centered competency-based practices revealed learning environments that use different competency definitions to guide learning expectations, grading approaches, school scheduling approaches, learning and graduation pathways, assessment systems, classroom-based applications of learning, and opportunities for learning beyond the classroom. Many of these same practices were closely intertwined with personalized, project-based, and other learner-centered approaches. As a result, it was not immediately clear which aspects of the learning model were truly rooted in competency-based learning (CBL). To understand what set these models apart, we had to examine the underlying structures and intentions of CBL more closely.

This study asked a more fundamental question: **What does CBL actually look like for learners, and what remains recognizable across high schools that put it into practice differently?** To explore the answer, we studied ten public high schools that reflected substantially different approaches to CBL and served learners across varied geographic, community, and demographic contexts. That diversity mattered. If similar competency-based learner experiences remained visible across substantially different settings, those experiences could offer the field shared language that was not tied to one particular type of school or community. Across those differences, we found a consistent throughline: **what learners demonstrated could influence what learning happened next, feedback and revision could strengthen their evidence of learning, and meaningful applied work could count as evidence of competency.** That finding gives the field clearer language for recognizing CBL across different models and understanding the conditions that make those experiences possible.

Academic data added another layer of context and insight to the study. Funders and policymakers have asked how learners who experience CBL perform on traditional assessment measures.

¹ National Academies of Sciences, Engineering, and Medicine (2018); Cantor et al. (2019); Bronfenbrenner and Morris (2006); Nasir et al. (2020).

² Bloom (1968); National Academies of Sciences, Engineering, and Medicine (2018); Brown et al. (2014); Ericsson and Pool (2016); Deans for Impact (2025).

³ Darling-Hammond et al. (2020); Immordino-Yang and Damasio (2007); Osher et al. (2020); National Academies of Sciences, Engineering, and Medicine (2018).

Participating high schools documented success on a number of external measures, including state assessments, standardized measures such as STAR or MAP⁴, advanced coursework such as Advanced Placement and dual enrollment, graduation rates, and matriculation to meaningful postsecondary opportunities. At two high schools with available data, the study found that learners who performed strongly on competency-based measures also tended to perform strongly on benchmark and state assessments.⁵

This study offers a grounded look at how CBL unfolds in high schools where learners are achieving at high levels. In doing so, we gain a clearer sense of what CBL actually looks like for learners preparing for postsecondary life, and how it can open up new pathways for meaningful academic growth.

Competencies Expand How Learning Looks

At the center of CBL are competencies: defined, cross-cutting academic and skills-based expectations for what learners are working toward and expected to demonstrate. The field has long wrestled with the relationship between academic knowledge and the development of practical skills. In a robust CBL environment, both must be present and valued. The challenge lies in ensuring that neither is sacrificed for the other, but that they work together to support meaningful learning and long-term success.

Competency development can become visible in many forms, both within and beyond the classroom. Learners demonstrate what they know and can do in relation to competencies not only through assessments and classroom projects, but also through experiences that connect them to the wider world, such as internships and work-based learning. These varied demonstrations can offer a fuller picture of learning in action. Looking at how learners encounter competencies and develop evidence across these experiences gives the field a common point of reference across models that may otherwise look very different.

That broader view of learning, however, still has to operate within a secondary education system organized around longstanding expectations for how progress and success are documented. Even as high schools move away from one-size-fits-all approaches, learners continue to earn credits, complete assessments used for accountability, meet established graduation requirements, and build academic records shaped by conventional reporting systems. CBL therefore asks high schools not only to create richer opportunities for learners to develop and demonstrate competency, but also to determine how that learning can remain visible and meaningful within the systems that surround it.

⁴ STAR and NWEA MAP are nationally normed progress monitoring assessments for English Language Arts and mathematics.

⁵ The analysis did not account for prior academic performance or learner demographic characteristics, which may influence outcomes.

Different CBL Models Reveal Shared Learner Experiences

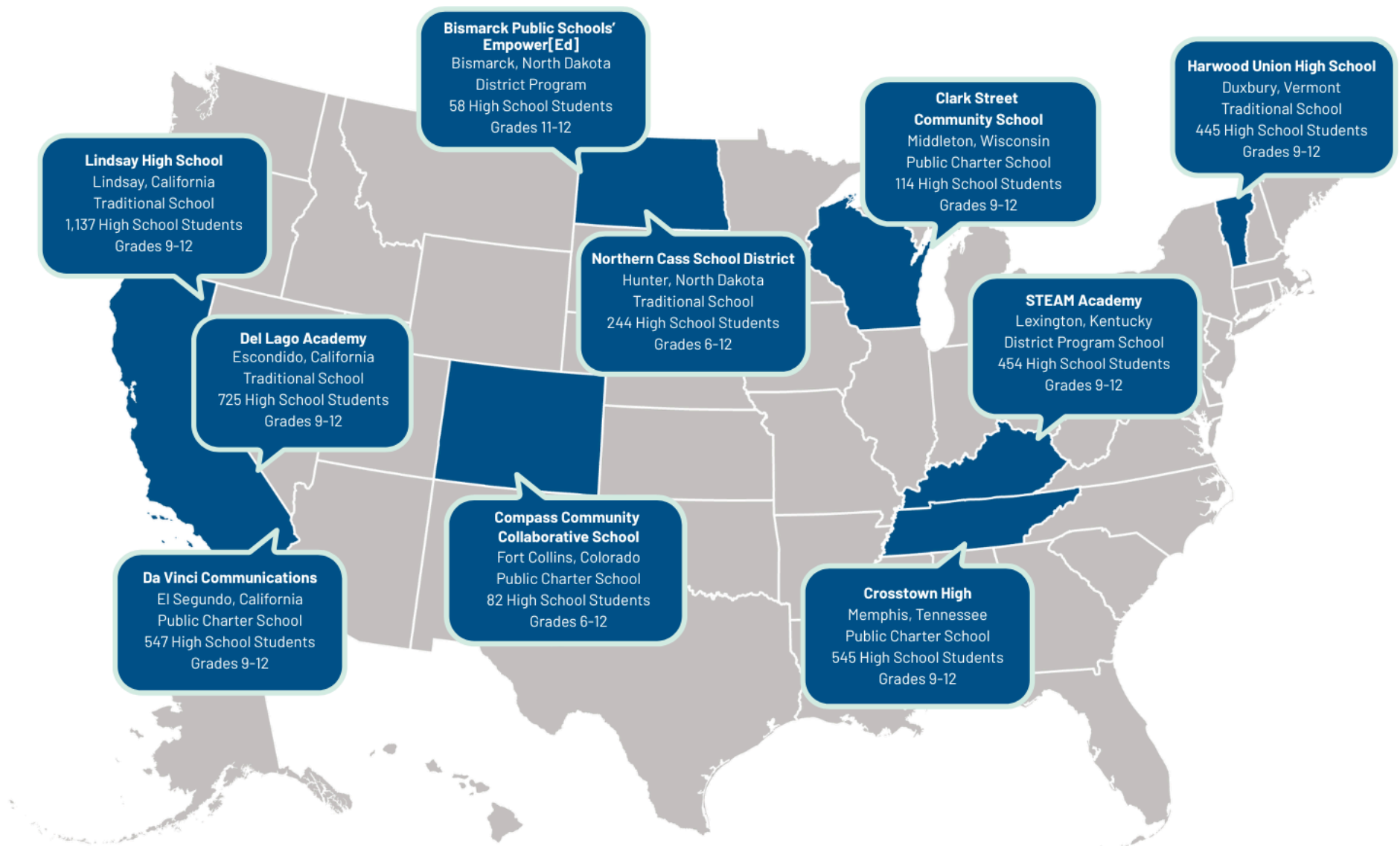
Competencies are foundational to CBL, but **the presence of competencies alone does not tell us how competency-based learning is experienced by learners.** To look beyond program labels and structures, FullScale entered the study with ten markers of competency-based learner experience, identified from existing research and field guidance and refined through researcher and advisory review.⁶ A *marker* describes something that may become visible when competencies meaningfully shape learning—how learners understand expectations, engage with evidence, use feedback, demonstrate competency, or move through learning. **Together, the ten provided a common learner-level lens for examining CBL across different high school models.**

The study intentionally examined those markers across high schools that put CBL into practice in substantially different ways. FullScale identified potential participants through a [national landscape scan](#) of competency-based secondary-school models and conversations with high school and system leaders. Final selection considered whether each high school had a substantive competency-based approach, whether evidence about learner success was available, and whether the overall group represented meaningful variation in context and model design. The resulting ten high schools, located in seven states, used different competency structures, schedules, grading approaches, pathways, and forms of applied learning.

That variation was central to the study. The goal was not to create a nationally representative sample or identify one preferred version of CBL. It was to ask whether similar learner experiences would remain recognizable across different approaches to CBL adoption, students served, and governance among the high schools. If they did, those experiences could provide the field with shared language for seeing CBL without tying it to one particular school model.

⁶ A full description of the marker development process is found in *Appendix B*. FullScale would like to thank the members of the CBL Advisory Board, who have provided invaluable feedback and guidance during this study. They include: Michael Crawford, America Succeeds; Bree Dusseault, CRPE; Jason Glass, Palo Alto Unified School District; Jillian Juman, Envision Education; Babak Mostaghimi, Learner Studio; Nikolas Namba, Transcend; Lillian Pace, KnowledgeWorks; Elena Silva, Learning Policy Institute; Laura Slover, Skills for the Future; Brook Stafford-Brizard.

Figure 1. Map of 10 Participating High Schools



No Single Pathway to Adopting CBL

The ten high schools that participated in our study each took a different route toward adopting CBL. Some were designed around competency-based principles from their inception. Others spent years incorporating those practices into established courses, schedules, grading systems, and graduation requirements. **What connected them was an intentional effort to organize meaningful dimensions of learning around what learners demonstrated in relation to defined competencies.**

Even the specific competencies took different forms. At some high schools, academic standards functioned as both academic and skills-based competencies because they organized expectations, assessment, and demonstrations of proficiency across contexts. At others, broader competencies such as *communication, critical thinking, leadership, or global citizenship* extended across disciplines and experiences.

Several high schools combined academic expectations with a Portrait of a Learner, a shared description of the knowledge, skills, and dispositions learners were expected to develop before graduation.

Table 1 shows that diversity more concretely by describing each high school's competency framework and the learner-facing features through which competencies became visible.

Table 1. Ten High Schools, Different Paths to Competency-Based Learning⁷

High School	Competency Framework	Learner-Facing Features
Empower[Ed], North Dakota	Academic standards used as competencies	Learners use evidence of mastery to prioritize work, progress at different rates, and demonstrate learning through interest- and career-connected projects.
Clark Street Community School, Wisconsin	Academic and transferable competencies	Learners develop <i>Artifacts</i> as competency evidence, use common performance criteria, and track progress through learning records and portfolios.
Compass Community Collaborative School, Colorado	Thirty-two academic and transferable competencies	Personalized pathways, portfolios, exhibitions, and community-connected experiences allow learners to develop and demonstrate competencies through varied forms of evidence.

⁷ These descriptions summarize the models examined in the study; they are not classifications of implementation.

High School	Competency Framework	Learner-Facing Features
Crosstown High, Tennessee	XQ Competencies	Competencies are incorporated into coursework and annual Presentations of Learning, where learners select evidence and explain their academic and transferable learning.
Da Vinci Communications, California	Course-level <i>Essential Knowledge & Skills</i> and broader <i>Habits of Mind</i>	Projects, rubrics, feedback cycles, presentations, and career-connected experiences help learners develop, apply, and explain competency evidence.
Del Lago Academy, California	Course-level competencies, called <i>Essential Course Competencies</i>	Performance tasks, revision, reassessment, interdisciplinary projects, internships, and presentations provide multiple opportunities to demonstrate proficiency.
Harwood Union High School, Vermont	Six schoolwide competencies, called <i>Learning Expectations</i>	Shared proficiency criteria connect learning across courses while operating within conventional credits, transcripts, and graduation requirements.
Lindsay High School, California	Course learning and mastery targets alongside broader <i>Lifelong Learning Standards</i>	A learning-management platform, flexible use of time, feedback, intervention, and career pathways support learners in understanding and progressing toward competency.
Northern Cass, North Dakota	Academic standards and a <i>Portrait of a Learner</i> describing broader competencies	Proficiency scales, personalized progressions, interdisciplinary learning, and work-based experiences provide opportunities to develop and demonstrate academic and transferable competencies.
STEAM Academy, Kentucky	Six shared competencies, called <i>STEAM Habits</i> , alongside academic standards	Projects, feedback, and Defenses of Learning connect academic expectations and broader competencies across the high school experience.

The differences across the ten high schools are what make the cross-site findings useful. Any learner experience that remained recognizable across them had to travel across different competency structures, instructional approaches, and systems rather than depend on one particular architecture.

Applying a Common Lens to Understand Different Models

The participating high schools defined and organized competencies in different ways, creating a need for a common lens on what those differences meant for learners in practice. Visible practices such as flexible pacing, projects, portfolios, or mastery-based grading could not serve as that common point. **Those practices may be important, but they do not tell us on their own how competency is experienced by learners.**

The ten CBL markers identified during the initial stage of this study allowed us to look across substantially different high school models without requiring those models to use the same implementation approach. A learner might encounter competency through a proficiency scale in one high school, a project or exhibition in another, or a learning-management platform somewhere else. The markers offered a common language for examining what those different structures meant for learners. In the sections that follow, we clarify what these markers represent, consider their dimensions, and reflect on how they should be understood in the broader context.

The ten markers were organized within two broad dimensions (Figure 2). The *Learning Design* dimension focuses on how competencies shape what learners understand about their learning—what they are working toward, how expectations develop over time, how they see their progress, and how demonstrated proficiency may influence what comes next. The *Assessment* dimension focuses on how evidence of competency is developed and used—where evidence can come from, how learners strengthen it through feedback and revision, how they apply competencies in meaningful contexts, and how evidence is interpreted and judged.



Figure 2. Ten Markers of Competency-Based Learning Applied in Study⁸

Learning Design	Assessment
Clear Articulation The extent to which learners organize and make meaning of school and lived experiences around cross-cutting competencies.	Demonstration Location The extent to which learners identify, collect, and submit evidence of competency from school and lived experiences.
Developmental Progression The extent to which learners recognize and engage with increasing complexity and depth within a competency over time.	Clear Calibration The extent to which competency evidence is interpreted using shared criteria intended to support accurate and consistent judgments across assessors.
Connection to Learning Targets The extent to which learners connect content- or skill-specific learning targets to broader competency development.	Iterative Purpose The extent to which learners seek and use adult and peer feedback, revise through multiple drafts or attempts, and request targeted feedback to strengthen evidence toward mastery.
Transparency The extent to which learners can access and use records of competency evidence to understand their progress and what remains.	Rigorous Application of Learning The extent to which learners produce authentic, contextually meaningful work that serves as evidence of mastery aligned with competency criteria.
Universal Application of Expectations The extent to which learners recognize and use common competency-based performance criteria across different work, educators, or contexts.	
Mastery-Based Progression The extent to which demonstrated proficiency influences when learners move forward and where they continue working.	

⁸ The *Learning Design* and *Assessment* dimensions organize the markers but do not imply a sequence or hierarchy.

Different Data Sources Reveal Different Parts of the Learner Experience

Once the study had a common set of learner experiences to look for, the next challenge was seeing them clearly. **No single data source could capture the whole of a learner's experience with CBL.** Some experiences unfold in a classroom moment; others develop across repeated attempts, courses, projects, or years.

- Student and teacher surveys offered insight into schoolwide perceptions of learner experiences.
- Classroom observations provided snapshots of what learners and educators were doing within the study's ten competency-based learning environments.
- Learner focus groups invited young people to share their stories and perspectives, offering us a deeper understanding of how learning unfolds over time and in different environments.
- High school materials and interviews helped researchers understand the models surrounding those experiences, while academic records and policy data provided context about outcomes and the systems in which those models operated.

These data sources were designed to complement one another in providing a fuller picture of CBL experiences. We expected that an observation might capture a learner revising one piece of work, for example, while a focus-group account could reveal that the work developed through several rounds of feedback over multiple weeks. Similarly, a survey could show that learners broadly reported an experience without explaining how it unfolded in a particular setting.

That difference in methods for gathering information became important to the study's findings. The strongest cross-site patterns that emerged were not experiences that looked identical in application across surveys, focus groups, and observations. The strongest patterns were those

The Study at a Glance

10 participating high schools
1,350 student survey responses
141 teacher survey responses
49 classroom visits / **97** observer records
105 learners participating in focus groups
17 policy conversations involving **45** participants
High school materials, policy documents, and available learner-outcome evidence

The study purposefully included high schools with substantive competency-based approaches and available evidence of learner success while representing different models and contexts. It was designed to describe learner experiences and the conditions surrounding them, not to determine that CBL caused learner outcomes or identify one preferred implementation model.

related to experiences that remained recognizable and distinct from others when viewed through these different lenses.⁹

Policy Context in This Report

Each participating high school operated within a distinct set of state and local policy conditions that shaped how competency-based learning could be designed and experienced by learners. FullScale conducted interviews and focus groups with school, system, and policy leaders to understand how those conditions influenced implementation in practice. Brief **Policy Context** callouts throughout the findings highlight where policies governing areas such as credit, grading, reassessment, staffing, applied learning, and reporting intersect with the learner experiences documented in the study.

Three Common Learner Experiences Across Ten Different Models

Across the ten high schools, three recognizable and distinct learner experiences emerged the most prevalently across settings and data sources:

- Learners encountered situations in which **demonstrated proficiency influenced when learners moved forward and what learning happened next** (*Mastery-Based Progression*).
- Learners used **feedback, revision, and additional attempts to strengthen evidence toward mastery** (*Iterative Purpose*).
- Learners **applied learning through meaningful work that could serve as evidence of mastery against defined competency expectations** (*Rigorous Application of Learning*).

These experiences remained recognizable and distinct across high schools with substantially different CBL models, **but they seemed to differ in meaningful ways**. To understand those differences, the study looked at three aspects of the learner experience: **the role competency played, the degree of learner ownership, and the scope of the experience across tasks, courses, or settings**. Together, these dimensions formed the *CBL Continuum* (Table 2), a way of describing variation within the same marker rather than rating learners, classrooms, or high schools.¹⁰

⁹ Technical details on measure development, reliability, observation agreement and dependability, qualitative coding, and analytic procedures are provided in *Appendices A-H*.

¹⁰ The CBL Continuum and its three dimensions were established before analysis of the study data. The continuum was applied to individual observation and focus-group records rather than to learners, educators, classrooms, or high schools as a whole. It should be interpreted as a descriptive lens for variation, not as a validated developmental sequence or measure of quality.

Table 2. CBL Continuum

Dimension	Pre-Emergent	Baseline	Maturing	Advanced
Competencies	Implicit	Explicitly referenced	Applied with support	Independently integrated
Ownership	Adult-directed	Adult-supported	Shared ownership	Learner-directed
Scope	Single task	Single course	Multiple experiences	Whole learning ecosystem

The examples that follow show both parts of the finding: what remained recognizable across models and how the same learner experience could take different forms.

Mastery-Based Progression: Competency Evidence Shapes What Learning Happens Next

Progression describes the extent to which demonstrated proficiency influences when learners move forward and where they continue working.

For Mastery-Based Progression, flexible pacing is not the defining feature. What matters is whether demonstrated mastery influences the learner's path. For example, during observations at STEAM Academy, pairs of learners were designing cardboard chairs capable of supporting an adult. Before beginning full-size construction, each pair built, tested, and refined a smaller model. During our observation, we saw some pairs measuring or troubleshooting their design. Other pairs, who had demonstrated readiness, were gathering materials to begin building their full-size chairs.

The visible difference in these experiences was the pace at which students worked on this project. The more important difference was **why learners were at different points.**

LEARNER VOICE

Time to Demonstrate Learning

At Da Vinci Communications, course-level competencies are called Essential Knowledge & Skills (EKS). One learner described the connection between time to understand the material and progression:

"I get actual time to learn and understand what I'm actually learning, and I just move to the next EKS, which is what I really loved about the school, because they give you time, understand what you know, to actually explain what you know, and actually show what you know."

Evidence from their models, connected to the project's engineering competencies, informed readiness for the next stage. Moving forward involved more than completing the same amount of work on the same timeline.

Mastery-Based Progression could still take different forms across learning experiences and school contexts. Progression might be guided primarily by an educator within a particular task or project, or learners might take greater ownership for monitoring their progress toward mastery and determining what they are ready to take on next. The experience can also range from progression within a single learning activity to movement across courses or broader learning pathways.

Across various data sources, learners described continuing to work where mastery had not been demonstrated, receiving additional support, and moving into subsequent learning once their evidence was sufficient.

Policy Context

Credit, Course Completion, and Graduation

When demonstrated proficiency influences how learners progress, high schools have to reconcile competency-based progressions with local systems organized around semesters, course-completion dates, reporting periods, and progress-monitoring expectations, as well as state policies governing credit, graduation, instructional time, and assessment.

These requirements create disconnects when a learner's readiness does not align with the system's calendar. A learner who needs additional time may continue developing and documenting evidence for first-semester coursework into the next term, while another may be ready to demonstrate learning before a course or reporting period ends. State assessment schedules may require learners to test at a prescribed time even when their readiness suggests an earlier or later opportunity would better align with their progression. **The policy challenge is not whether competency-based progression is allowed, but how credit, monitoring, and assessment systems can represent learning that does not unfold on a uniform timeline.**

Iterative Purpose: Feedback Strengthens Evidence of Competency

Iterative Purpose describes learners' use of adult and peer feedback, revision, and additional attempts to strengthen evidence toward competency mastery.

A Da Vinci Communications journalism course illustrates what this looked like in practice. Learners experimented with editing choices, sought feedback from peers and educators, and decided how to strengthen their documentary before submitting a final version. The course used a shared revision vocabulary—“*rough cut*,” “*fine cut*,” and “*final cut*”—while keeping communication competencies as a reference point for those decisions.

Feedback did not arrive only after the work was finished. It helped learners shape their evidence before the final product, with competency expectations providing a reference point for deciding what needed improvement.

Learner accounts from Del Lago Academy speak to a similar experience. Participants described tutorials, revision, and reassessment after receiving a *No Mark*, which indicated that the available evidence had not yet demonstrated mastery. They also described seeking feedback during everyday work.

Iteration also differed in who drove the process. In some experiences, educators identified what needed to change and guided learners through revision toward a competency expectation. In others, learners took greater responsibility for seeking feedback, interpreting it against expectations, and deciding what to revise. The Del Lago learner’s description of having to advocate for feedback illustrates this more learner-directed form. Both reflect *Iterative Purpose*; what varies is the degree of learner ownership.

Iterative Purpose was one of the strongest reported experiences in both the student and teacher surveys and appeared frequently in classroom observations. Across these sources, assessment could provide information learners used to improve—not only a judgment of competency proficiency.

Feedback and revision are not unique to CBL; what matters is if learners can act on feedback and produce stronger evidence that remains consequential to a competency assessment.

LEARNER VOICE

Seeking Feedback

One learner at Del Lago Academy described the role learners sometimes played in initiating that process:

“I think it's definitely you kind of have to advocate for yourself, and you know, is this fine? What can I do to improve? ... Teachers are very willing to give feedback, you just kind of have to also seek it for yourself.”

Policy Context

Grading and Reassessment

When feedback and additional attempts allow learners to produce stronger evidence of competency, grading systems have to determine **which evidence ultimately represents what the learner knows and can do**. Local reassessment and grading policies shape whether later demonstrations can replace or update earlier performance, how multiple attempts contribute to a course grade, and how competency judgments are translated into GPA, transcripts, and credit.

This creates a different systems challenge from progression. **An instructional model may treat revision as continued learning, while the grading system can still preserve earlier performance in ways that obscure what the learner ultimately demonstrated.** How schools record reassessment therefore affects whether iteration changes not only the learner's work, but the formal academic record used to communicate that learning.

Rigorous Application of Learning: Applied Work Becomes Evidence of Competency

Rigorous Application of Learning describes authentic, contextually meaningful work that serves as evidence of mastery in relation to defined competency expectations.

As an example, at Empower[Ed], a learner described working with a local historical society to produce recordings of scripts from the 1930s, detailing Bismarck history, for middle school students. The work involved considering historical voice, practicing voice acting, and connecting the final performance to English standards and communication competencies. The audience gave the work a purpose beyond submission to an educator; the competency connection established what learning the work could demonstrate.

The example makes an important distinction visible. The study was not simply documenting projects with real-world themes. It examined whether learners applied relevant knowledge and skills through work that could be interpreted against competency expectations.

LEARNER VOICE

Connecting a Public Product to Learning

In a discussion of community projects, an Empower[Ed] learner described considering both the partner's expectations and the educator's assessment:

"It really depends on what standard wanted to get for it, because then you just talk to your teacher, but demonstrating the proficiency through the projects, you talk to the community partner you're doing the project for. You get an idea of what they want, and you make sure that they're okay with the finished result. And then you go talk to your teacher, compare and contrast [how it will help you demonstrate your standards]."

Rigorous Application could also take different forms along the continuum. In the learner experience at Empower[Ed], competencies explicitly shaped how the work was interpreted, the learner navigated expectations from both the community partner and educator, and the experience extended beyond the classroom. Other examples were more bounded or adult-directed. **What remained consistent was that meaningful work became evidence of defined competency.**

Rigorous Application of Learning was especially visible in learner focus groups and classroom observations. Longer learner accounts could make the full arc of applied work visible in ways a brief observation or survey item could not.

Across the participating high schools, applied learning took many forms, from classroom projects and community-connected work to internships and other work-based learning experiences. Those experiences are not inherently competency-based. **The shift occurs when meaningful applied work becomes evidence of defined competency.**

Policy Context

Recognized Applied Learning

When projects, internships, or community-based experiences generate evidence of competency, leaders have to determine how that learning becomes formally recognized.

Work-based learning, alternative-credit, licensure and assignment, course-coding, and transcript policies can shape what evidence counts, who can certify it, and how it enters the academic record.

The challenge is not simply whether applied learning is permitted, but whether demonstrated learning from those experiences can be captured in formal records, translated into credit, and counted toward learners' progress through high school.

Across all three recurring markers, the same learner experience could take different forms.

Competencies could play an implicit to explicit role, learners could exercise different degrees of ownership, and experiences could range from a single task to learning that extended across courses, settings, or time. These differences did not make one experience more or less competency-based; they helped describe how a shared marker took shape within different high school models. **The value of the continuum is therefore not to rank implementation, but to make meaningful variation visible within experiences the high schools shared.¹¹**

¹¹ Detailed distributions of observation and focus-group records across the CBL Continuum are provided in Appendix B.

LEARNER VOICE

One Moment and a Longer Account

At Clark Street Community School, one learner described approximately 12 weeks researching a local opera house, consulting community sources, and creating models of its visual history:

“It was probably 12 weeks I spent doing research on the Opera House, because I go to a lot of shows there. I couldn’t drive at the time, so my mom just drove me back and forth, twice a week to the Historical Society. We met with people at the Opera House, went to the archives, and that was all through Project Block. Then I was able to expand it outside of the classroom and build models to represent this visual history. They’re on display there, which is pretty sweet.”

A brief classroom visit might have captured only one stage of that work. The learner’s account made the longer experience visible.

The Full Set of Markers Strengthens How Competencies Shape the Learner Experience

The three recurring markers were the clearest cross-site patterns in the study, but they were not the only competency-based learner experiences documented. **Evidence related to all ten markers appeared to some degree across the participating high schools.** What distinguished *Mastery-Based Progression*, *Iterative Purpose*, and *Rigorous Application of Learning* was the strength and consistency of the pattern across different high schools and data sources. The other seven markers captured additional ways competencies shaped learners’ experiences: making sense of broader competencies (*Clear Articulation*), seeing how competencies develop over time (*Developmental Progression*), connecting everyday learning to broader expectations (*Connection to Learning Targets*), tracking progress and what remains (*Transparency*), using common expectations across different work and settings (*Universal Application of Expectations*), drawing evidence from varied experiences (*Demonstration Location*), and understanding how evidence will be judged (*Clear Calibration*).

The same marker could also be supported through different approaches to CBL. *Transparency*, for example, focuses on whether learners can access and use information about their competency progress. At Lindsay High School, learners used a digital learning-management platform to see targets, proficiency ratings, completed work, and remaining assignments. At Clark Street, learners used learning records and portfolios to track evidence against shared competencies. The mechanisms differed, but the learner experience was similar: learners could see what they had demonstrated and what remained.

Some markers also require a longer window to become visible. Seeing how a competency develops in complexity over time (*Developmental Progression*) may require looking across several years, while evidence developed through internships, community settings, or other experiences beyond the classroom (*Demonstration Location*) may not appear during a single classroom visit. Different data sources therefore helped reveal different parts of these experiences.

The three recurring markers provide the clearest cross-site signals of where competency became visible in learners' experiences. The full set of markers helps us see the unique implementation of CBL by capturing a broader range of ways competencies can shape what learners understand, do, demonstrate, and experience.

The Role Competency-Based Learning Plays Within Other Learner-Centered Approaches

Many high schools intentionally combine multiple learner-centered approaches—such as project-based learning, personalization, and flexible pacing—within the same learning model. That makes it especially important to understand what competency contributes within that mix.

Learners across the ten high schools worked on projects, exercised choice, pursued interests, moved at different rates, and connected learning with experiences beyond the classroom. Those practices also appear in project-based learning (PBL), personalization, and other learner-centered approaches. Responses from the Student Survey helped us better understand the interplay between these approaches. We found that, within the high schools in our study, learners who reported more competency-based experiences also tended to report more project-based and personalized learning experiences.

The overlap is real, but the approaches are not the same:

- **PBL** organizes learning through inquiry, disciplinary work, projects, and products.
- **Personalization** is learning that is responsive to learners' needs, interests, goals, pace, and pathways.
- **CBL** uses explicit competencies to give those experiences a particular organizing logic. The competencies shape what learners are working toward, what can count as evidence, how that evidence is interpreted and strengthened, and how demonstrated proficiency may matter for what happens next.

The visible practice alone does not make an experience competency-based. A learner can choose how to demonstrate learning, complete a meaningful project, or move at an individualized pace without competencies necessarily shaping the experience. **The competency-based dimension lies in the connection between those practices and defined expectations for what learners are expected to know and be able to demonstrate.**

The three recurring markers make this distinction particularly visible.

- Flexible pacing becomes *Mastery-Based Progression* when demonstrated proficiency influences what happens next.
- Feedback and revision become *Iterative Purpose* when learners use them to strengthen evidence that matters to a competency judgment.
- Applied work becomes *Rigorous Application of Learning* when the work itself demonstrates defined competency expectations.

LEARNER VOICE

When Learner-Centered Approaches Overlap

At Compass Community Collaborative School, one learner described a science project that brought personalization, project-based learning, and competency together:

"I remember I took a science class one year, and for my final project, I did a creative writing piece where, in the piece, I explained in detail how scaling works in science, and I got full credits for it."

What distinguishes CBL is not the practice itself, but the role competencies play in shaping what learners are expected to demonstrate and how that evidence is used. That organizing role also creates a learning environment in which improvement can remain consequential over time.

Competency-Based Experiences Were More Closely Related to Growth-Oriented Beliefs About Learning

CBL treats mastery as something learners can continue developing, not simply something judged at a single point in time. When learners can use feedback, revise their work, make additional attempts, and continue developing mastery, growth and improvement become part of how learning happens.

The Student Survey results reflected that orientation. Learners who reported more competency-based experiences also tended to place greater emphasis on developing understanding, building skills, improving through effort, and becoming more capable over time.

In other words, learners reporting more CBL experiences also tended to hold a more growth-oriented view of learning.¹²

¹² The Student Survey measured *Mastery Goal Orientation*, which emphasizes developing understanding, skills, and competence through learning and effort, and *Performance Goal Orientation*, which emphasizes demonstrating ability relative to others. Reported CBL experiences were consistently more closely related to Mastery Goal Orientation across the markers examined. Although related to the concept of "growth mindset," Mastery Goal Orientation is a distinct construct. Because the measures were collected at the same point in time, these findings describe associations and do not establish that CBL caused learners to develop particular beliefs. Full statistical results appear in *Appendix D*.

LEARNER VOICE

Growth Rather Than Comparison

A Northern Cass learner described educators emphasizing progress relative to the learner's own earlier performance:

"I think a lot of teachers with us have emphasized this idea of not necessarily comparing yourself to what other people are doing, and more so comparing yourself to you from last year. I think a big part of it is that growth, rather than competing against others for a better grade."

High Schools Demonstrate Strong Performance on Conventional Success Measures

While learner-level findings reveal how CBL takes shape across high schools with varied approaches, it is important to recognize that these schools remain deeply embedded within the larger secondary education system. Students continue to navigate familiar structures: earning credits, completing external assessments, building transcripts, and preparing for the next steps after graduation, whether that means college or entering the workforce.

Conventional academic outcomes measures therefore remain important—not because they define whether a high school is competency-based, but because they remain some of the primary ways families, states, postsecondary institutions, and the public understand learner success.

The study examined available outcomes data to understand what academic success was documented in the same settings where competency-based learner experiences were visible.¹³ The participating high schools tracked different measures of learner success, so the study did not create a single cross-site scorecard. Instead, we reported verified outcome data that were most relevant to each high school's context. Table 3 highlights several of the traditional outcome measures documented across the participating high schools, with reference to a local, contextual comparison.

¹³ Academic outcomes are descriptive and contextual. Participating high schools were purposefully selected in part for available evidence of learner success; external comparison groups were not statistically matched. Most outcome records describe the 2024–25 school year, while primary learner-experience data were collected in spring 2026, preventing direct analysis of whether the measured learner experiences produced later outcomes. Comparisons provide descriptive context rather than matched comparisons or estimates of impact. The Crosstown and Memphis-Shelby County Schools charter-school results are from different reporting years. Methodology is provided in *Appendix H*. Full definitions, denominators, reporting periods, and additional outcomes appear in the individual high school case studies (November 2026).

Table 3. Illustrative Conventional Academic Outcome Measures

High School	Outcome Measure	Participating High School Outcome	Contextual Comparison	What the Comparison Shows
Clark Street Community School	State ELA assessment	82.5% of tested learners reached <i>Meeting or Advanced</i>	75.6% at Middleton High School	Among tested learners, Clark Street's ELA performance was higher than the local comprehensive high school comparison.
Crosstown High	AP exam performance	53.2% of 447 AP exams earned a score of 3 or higher	26.3% among Memphis-Shelby County Schools charter schools	Crosstown combined broad AP access with a higher qualifying-score rate than the district's charter-school reference group.
Da Vinci Communications	College/ Career Indicator: Prepared	83.6% of graduates	60.3% at CHAMPS Charter High School of the Arts	A larger share of Da Vinci graduates met California's multidimensional college- and career-readiness criteria than at the comparison high school.
Del Lago Academy	Four-year graduation rate	97.8%	86.8% across Escondido Union High School District	Del Lago's graduation rate was higher than the district reference point.

Local Competency Evidence Relates to Conventional Academic Outcomes Measures

At two high schools, the available records allowed the study to look more closely at the relationship between local competency evidence and academic performance recognized outside the local system.

At Del Lago Academy, course grades reflected demonstrated proficiency across course competencies. Learners with stronger competency-based GPAs also tended to perform more strongly on Grade 11 state English language arts and mathematics assessments.

At Northern Cass, the local competency measure captured the proportion of mathematics standards on which learners had demonstrated proficiency. Learners demonstrating proficiency across more standards, leveraged as competencies, also tended to perform more strongly on STAR Mathematics, the ACT, and, among Grade 10 learners, the state mathematics assessment.

Table 4 summarizes the two local measures, the external measures examined, and the common directional pattern.¹⁴

Table 4. Local Competency Evidence Relates to External Academic Performance

High School	What the Local Competency System Captured	External Measures Examined	What We Found
Del Lago Academy	Demonstrated proficiency across course competencies, reflected in competency-based GPA	Grade 11 state ELA and mathematics assessments	Learners with stronger performance in the local competency system also tended to perform more strongly on both state assessments.
Northern Cass	Proportion of mathematics standards on which learners demonstrated proficiency	STAR Mathematics, ACT composite, and Grade 10 state mathematics assessment	Learners demonstrating proficiency across more mathematics standards also tended to perform more strongly on external academic measures.

¹⁴ Analyses were limited to learners with records available on both local competency and external academic measures. The relationships describe association, not causation, interchangeability, or added predictive value. Methodologies appear in the Appendix H.

The measures were not assessing learning in identical ways. External assessments capture performance under particular testing conditions and at particular points in time. Local competency systems may draw on repeated demonstrations, coursework, reassessment, and evidence accumulated over a longer period.

The finding is not that local competency measures simply reproduced external scores. It is that learners demonstrating stronger competency performance locally also tended to demonstrate stronger academic performance externally.

That relationship matters because local competency measures and external academic measures capture different dimensions of learner performance. Competency systems can make visible a broader range of knowledge, skills, and demonstrations developed over time, while external measures provide signals that are already widely recognized across schools and systems. The positive relationships at Del Lago Academy and Northern Cass suggest that learners can build and demonstrate a broader set of competencies while also performing successfully on conventional academic measures.

Policy Context

When Competency Progress Does Not Fit the System

Across participating high schools, leaders described a practical mismatch between competency-based learning and the systems used to record academic progress. Student information systems and learning-management systems were often designed around assignments, courses, grades, credits, and reporting periods—not around competency development extending across courses, repeated demonstrations, or growth over time.

As a result, high schools sometimes relied on separate spreadsheets, portfolios, specialized platforms, or locally developed processes to preserve information about competency progress. Others adapted systems that were not designed to represent revision, reassessment, or evidence accumulating across multiple experiences.

That fragmentation made it difficult to connect detailed competency records with assessment results, grades, credits, graduation data, and other information maintained in central reporting systems. **When competency progress does not flow into the systems district and state leaders use to understand learner performance, the relationship between competency development and conventional outcomes can remain largely invisible to the people making system-level decisions.**

Competency Evidence Must Travel Across Systems

The relationships documented at Del Lago Academy and Northern Cass were possible to examine because local competency records could be connected with external academic data at the learner level. Across other high schools, that connection was much harder to make.

The challenge was not necessarily that competency information did not exist. High schools often had detailed records of learner progress, but those records could be stored in different platforms or formats from the grades, assessment results, credits, and other information used for formal academic reporting. Local competency measures also differed in what they captured and how they were recorded. **As a result, schools could have rich information about competency development without an easy way to examine it alongside other indicators of learner progress.**

That matters for building a stronger evidence base. If the field wants to understand how competency development connects with academic, postsecondary, and other outcomes, learner evidence needs to remain usable as it moves across courses, platforms, and reporting systems. Better connections would allow high schools, system leaders, and researchers to examine different dimensions of learner progress together without assuming that they measure the same thing.

The need, then, is not to replace conventional academic outcomes measures with competency-based outcomes measures. It is to make different forms of learner evidence connectable so the field can develop a fuller picture of what learners know, can do, and accomplish over time.

Policy Context

Records, Reporting, and Accountability

Policies governing grades, credits, course codes, transcripts, and state reporting shape how competency evidence enters the official academic record. When local competency-tracking systems and formal reporting systems do not align, high schools may have to translate detailed information about competency progress into conventional measures such as grades or credits, **flattening some of the nuance and meaning the original evidence was designed to preserve.**

Better alignment would make it easier for local leaders and policymakers to examine competency progress alongside assessment results, graduation, and other indicators already used for accountability and decision-making. **That connection is a basic condition for stronger evidence-building: different forms of learner evidence need to be linkable across systems without being collapsed into the same measure.**

Shared Language Creates a Foundation for a Stronger CBL Evidence Base

Across ten intentionally different high schools, CBL became more recognizable when the study looked beneath the model and closer to the learner experience. The commonality was not one schedule, grading system, project model, or technology platform. It was the role competencies could play in organizing learning and making evidence consequential. Because the participating high schools were purposefully selected rather than nationally representative, these findings should not be read as estimates of how often particular experiences occur across all CBL high schools. **Their value lies instead in identifying learner experiences that remained recognizable across substantially different approaches to putting CBL into practice.**

That gives the field something important to build from: **shared language that can travel across models without requiring those models to become the same.** Schools can use that language to examine whether the experiences they intend are reaching learners. Researchers can use it to investigate those experiences more consistently across settings and over time. System leaders can ask more precise questions about the conditions that allow competency evidence to matter and whether their data and reporting systems are capable of seeing it.

The study also creates a foundation for connecting learner experience with a broader picture of learner progress. The measures developed here are still evolving, and different data sources captured different dimensions and time horizons of the learner experience. The CBL Continuum remains a hypothesis for describing variation rather than a validated developmental sequence, and the available outcome and policy analyses provide descriptive and contextual insight rather than evidence of causal effects. Even so, the work moves the field beyond relying primarily on program labels or descriptions of model design. It creates an empirical starting point for refining how CBL is measured, examining how these experiences develop across settings and over time, and studying how competency evidence relates to academic, postsecondary, and other longer-term outcomes.

The significance of this study is therefore not that it identifies one final version of CBL. It gives the field a clearer, evidence-grounded way to see competency-based learning across models—and a shared foundation for building a deeper evidence base around what those experiences mean for learners.

What's Coming Next

This report is one part of a broader effort to make competency-based learning more visible, usable, and actionable for the field. Upcoming resources will build on these findings in different ways:

- **High School Case Studies** will take a closer look at each of the ten participating high schools, showing how competencies were organized locally, how the markers became visible in learners' experiences, and what available academic and outcome data add to the picture.
- **A Policy Report** will examine the state and local policy conditions surrounding these high schools, including how leaders navigated rules related to credit, grading, assessment, staffing, applied learning, and reporting to make CBL workable in practice.
- **A Practitioner Hub** will translate the study findings into accessible tools and resources that schools and systems can use to examine learner experiences, reflect on their own approaches, and support continuous improvement.

These resources will provide complementary ways to understand CBL: through the experiences of learners, the choices and conditions shaping implementation, and practical tools for strengthening the work.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- ASCD. (2007). *The learning compact redefined: A call to action*.
<https://library.ascd.org/m/21e2f544234c3e97/original/WCC-Learning-Compact.pdf>
- Big Picture Learning. (n.d.). *Who we are*. <https://www.bigpicture.org/whoweare>
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, Article 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Brennan, R. L. (2001). *Generalizability theory*. Springer.
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (6th ed., pp. 793–828). Wiley.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Brown, P. C., Roediger, H. L., III, & McDaniel, M. A. (2014). *Make it stick: The science of successful learning*. Belknap Press of Harvard University Press.
- Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2019). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science*, 23(4), 307–337.
- Carnegie Foundation for the Advancement of Teaching. (2025). *A research and development agenda for high school transformation*.
<https://www.carnegiefoundation.org/wp-content/uploads/2025/07/RD-Agenda.pdf>
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, 6(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1
- Cognia. (2023). *Effective learning environments observation tool, version 2.0: Technical brief*.
<https://www.cognia.org/wp-content/uploads/2023/01/eleot-Technical-Brief.pdf>

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, 24(2), 97–140.
- Deans for Impact. (2025). *The science of learning* (2nd ed.). <https://www.deansforimpact.org>
- Education Policy Innovation Collaborative. (2021). *Competency-based education survey instrument report*. Michigan State University.
- Education Reimagined. (2020). *A transformational vision for education in the U.S.* <https://education-reimagined.org/wp-content/uploads/2024/11/A-Transformational-Vision-for-Education-in-the-US.pdf>
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), 430–457.
- Ericsson, K. A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. Houghton Mifflin Harcourt.
- Evans, C. M., Landl, E., & Thompson, J. (2021). Making sense of K–12 competency-based education: A systematic literature review of implementation and outcomes research from 2000 to 2019. *Journal of Competency-Based Education*, 5(4), Article e01228. <https://doi.org/10.1002/cbe2.1228>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—Principles and practices. *Health Services Research*, 48(6, Pt. 2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- Germain, E., Oakes, J., & Maier, A. (2023). *Theory of action for community school transformation*. Learning Policy Institute. <https://learningpolicyinstitute.org/project/community-schools-forward>
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255–274. <https://doi.org/10.3102/01623737011003255>

- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238–247. <https://doi.org/10.1037/1040-3590.7.3.238>
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, 1(1), 3–10.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36(2), 109–133. <https://doi.org/10.1007/BF02291393>
- 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>
- LEAP Innovations. (2024). *LEAP learning framework for personalized learning: Overview*. <https://www.leapinnovations.org/llf>
- Levine, E., & Patrick, S. (2019). *What is competency-based education? An updated definition*. Aurora Institute. <https://aurora-institute.org/wp-content/uploads/what-is-competency-based-education-a-n-updated-definition-web.pdf>
- Martin, K. (2025). *A learner-centered ecosystem*. Learner-Centered Collaborative. https://learnercentered.org/wp-content/uploads/2025/06/a_learner-centered_ecosystem.pdf
- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(2), 13–17. <https://doi.org/10.3102/0013189X017002013>
- McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning*, 22(1), 65–72. [https://doi.org/10.1016/S0149-7189\(98\)00042-1](https://doi.org/10.1016/S0149-7189(98)00042-1)
- Midgley, C., Maehr, M. L., Huda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales*. University of Michigan.
- Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (Eds.). (2020). *Handbook of the cultural foundations of learning*. Routledge.

- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>
- Osher, D., Cantor, P., Berg, J., Steyer, L., & Rose, T. (2020). Drivers of human development: How relationships and context shape learning and development. *Applied Developmental Science*, 24(1), 6–36.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Parker, A., & Tritter, J. (2006). Focus group method and methodology: Current practice and recent debate. *International Journal of Research & Method in Education*, 29(1), 23–37. <https://doi.org/10.1080/01406720500537304>
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, 25(1), 6–14. <https://doi.org/10.1177/1094428120968614>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. SAGE Publications.
- Stafford-Brizard, K. B. (2016). *Building blocks for learning: A framework for comprehensive student development*. Turnaround for Children. <https://turnaroundusa.org/wp-content/uploads/2016/03/Turnaround-for-Children-Building-Blocks-for-Learningx-2.pdf>
- Sturgis, C., & Casey, K. (2018). *Quality principles for competency-based education*. iNACOL. <https://www.aurora-institute.org/wp-content/uploads/Quality-Principles-Book.pdf>
- The Learning Accelerator. (n.d.). *Powerful student learning*. <https://learningaccelerator.org/powerful-student-learning/>
- Transcend. (2025). *Leaps toward extraordinary learning for all*. https://transcendeducation.org/wp-content/uploads/2025/06/Leaps-2-Page-Overviews_v3_20250226.pdf

Wilson, V. (1997). Focus groups: A useful qualitative method for educational research? *British Educational Research Journal*, 23(2), 209–224. <https://doi.org/10.1080/0141192970230207>

Appendix A.

Study Design, Sample, and Data Collection

Study Purpose and Design

FullScale conducted a multi-site, mixed-methods study to examine how competency-based learning (CBL) becomes visible through learners' experiences across ten varied high schools. The study began with ten hypothesized markers of CBL and investigated whether those experiences appeared across different high school models, how learners and educators perceived and described them, and how their breadth varied within the available evidence.

The learner-experience study used a convergent mixed-methods design, supplemented by retrospective outcome analyses. Student and teacher surveys, classroom observations, learner focus groups, interviews, and high school materials provided complementary views of learner experience and were integrated during interpretation. This approach allowed the research team to examine the same broad phenomenon from multiple perspectives without assuming that any one source could fully represent the learner experience. Mixed-methods research is particularly useful when different sources illuminate complementary dimensions of a complex phenomenon and are intentionally brought together during interpretation (Fetters et al., 2013; Greene et al., 1989).

The study was exploratory, observational, and descriptive. It did not introduce an intervention, assign learners to different conditions, or include a comparison group of high schools not implementing CBL. Participating high schools were purposefully selected and do not constitute a nationally representative sample. Findings therefore describe patterns within these ten high schools and provide a foundation for hypotheses that can be tested and refined through future research; they do not estimate the prevalence or effects of CBL nationally.

Learner-experience data were collected primarily during spring 2026. The study was not longitudinal, and classroom observations captured 20 to 40 minutes of learning at one point in the high school year. Learner outcomes were drawn primarily from internal and public data describing the 2024–25 school year, the most recent complete year available for much of the analysis. These outcomes provide descriptive context for learner performance within the participating high schools; they were not temporally aligned outcomes linked to the spring 2026 learner-experience measures.

A parallel policy analysis examined the state and local conditions surrounding CBL implementation. Selected findings inform this report's interpretation of the authority, operating rules, and implementation support surrounding learners' experiences. A companion policy report presents the full cross-site analysis.

Research Questions

The study addressed three related areas of inquiry: the trustworthiness of the measures, the learner experiences visible across participating high schools, and the outcomes high schools identified and documented.

Measurement and Trustworthiness

- RQ1. What evidence supports interpretations and inferences drawn from the classroom observation protocol about observable markers of competency-based learning?
- RQ2. What evidence supports interpretations and inferences drawn from the student survey about learners' experiences of competency-based learning?
- RQ3. What evidence supports interpretations and inferences drawn from the teacher survey about learners' experiences of competency-based learning?

These questions treat validity as an evidence-based argument about the interpretations and uses a measure can support rather than as a fixed property of an instrument (Kane, 2013). Analyses examined item functioning, internal consistency, internal structure, relationships with theoretically related constructs, marker observability, inter-rater agreement, observer effects, and score dependability as appropriate to each measure.

Complete student-survey methods and findings appear in *Appendix C*. Teacher-survey analyses appear in *Appendix D*. Observation-protocol development, observability, inter-rater analyses, observer effects, and generalizability analyses appear in *Appendix E*.

Learner Experiences

- RQ4. Which of the ten hypothesized markers of competency-based learning were evident within and across participating high schools and evidence sources?
- RQ5. How did learners and educators perceive and describe these markers, and what variation in breadth was visible in scorable classroom-observation and learner focus-group evidence?

The complete marker framework and hypothesized breadth-of-experience continuum appear in *Appendix B*. The continuum was used to describe the breadth visible within individual observation records and learner accounts, including the role of competencies, learner ownership, and the scope of the experience. It was not used to rank high schools or produce a single high school-level score.

Learner Outcomes

- RQ6. What academic and broader learner outcomes did participating high schools identify as most closely connected to their learning models, and what did accessible 2024–25 data show about learner progress and performance?
- RQ7. How did participating high schools' outcomes compare descriptively with locally relevant high school, district, or state reference points?
- RQ8. Where compatible learner-level measures were available, what relationships appeared between local competency measures and conventional academic measures?

The study did not test relationships between learners' survey-reported CBL experiences and their academic outcomes. Analyses relating local competency evidence to conventional academic measures were limited to high schools with compatible learner-level records. Detailed outcome definitions and analytic procedures appear in *Appendix F*.

Policy Conditions

- RQ9. What state and local policy conditions characterize the environments in which these high-quality CBL schools operate?

The parallel policy inquiry examined how state and local policies enabled, constrained, or otherwise shaped CBL implementation across the participating sites. It considered what policy made possible, how local leaders interpreted and used available authority, where requirements created friction, and what support helped schools put their models into practice. This inquiry provided context for understanding learner experiences; it did not test whether particular policies produced stronger markers or learner outcomes.

Selecting Participating High Schools

Participating high schools were identified through a national landscape scan of competency-based secondary models. Researchers reviewed publicly available high school websites, reports, program descriptions, and related materials for evidence that high schools had intentionally organized meaningful dimensions of learning around defined competencies.

The research team then held informational conversations with high school and system leaders to clarify each model and determine whether participation was feasible. These conversations occurred before research data collection. They informed site selection but were not conducted using research protocols or analyzed as study evidence.

Final selection considered the substance of each high school's reported competency-based approach, available evidence of learner success, the feasibility of collecting multiple forms of evidence, and variation in high school context and model design. The study intentionally included high schools using different competency frameworks, schedules, grading systems, approaches to progression, and forms of applied learning. Purposeful sampling was appropriate because the aim was to study information-rich cases across varied contexts rather than estimate characteristics of a national population (Palinkas et al., 2015).

Because evidence of learner success informed site selection, the outcome findings should not be interpreted as representative of typical outcomes across all high schools implementing competency-based practices. Rather, the sample was designed to examine learner experience across a varied group of high schools with substantive competency-based approaches and evidence that learners could experience success within those settings.

The resulting sample included ten high schools in seven states. Some used academic standards as cross-cutting competencies. Others organized learning around high school-wide or cross-disciplinary competencies, and several combined academic expectations with a Portrait of a Graduate or another transferable-competency framework.

Table A1. Participating High Schools and Competency Frameworks

High school	Location	Competency Framework
Bismarck Public Schools' Empower[Ed]	Bismarck, North Dakota	Academic standards function as competencies and <i>Career Readiness Practices</i>
Clark Street Community School	Middleton, Wisconsin	<i>Elements</i>
Compass Community Collaborative School	Fort Collins, Colorado	Thirty-two academic and transferable competencies
Crosstown High	Memphis, Tennessee	<i>XQ Competencies</i>
Da Vinci Communications	El Segundo, California	Course-level <i>Essential Knowledge & Skills</i> and broader <i>Habits of Mind</i>
Del Lago Academy	Escondido, California	<i>Essential Course Competencies</i>
Harwood Union High School	Duxbury, Vermont	Six high school-wide <i>Learning Expectations</i>

Lindsay High School	Lindsay, California	Course learning and <i>Mastery Targets</i> alongside <i>Lifelong Learning Standards</i>
Northern Cass High School	Hunter, North Dakota	Academic standards and <i>Portrait of a Learner</i> competencies
STEAM Academy	Lexington, Kentucky	Six <i>STEAM Habits</i> alongside academic expectations

The sample was designed to make variation analytically useful. The participating high schools should not be interpreted as representing all high schools implementing competency-based practices, and their inclusion does not establish a national benchmark or comparison group.

Data Sources and Collection Procedures

The study combined student and teacher surveys, classroom observations, learner focus groups, teacher interviews, high school materials, supplemental learning settings, and learner outcomes. Each source contributed a different view of learner experience.

Table A2 summarizes the evidence collected and distinguishes the overall evidence collected from the units used in particular analyses. Because learners and educators could contribute to more than one evidence source, these counts should not be summed to produce an unduplicated study sample.

Table A2. Evidence Collected and Primary Analytic Samples

Evidence Source	Evidence Collected	Primary Analytic Sample or Unit	Role in the Study
Classroom observations	49 classroom visits; 97 observer records	Observer records and scorable marker evidence	Learner experiences visible during particular classroom episodes
Student surveys	1,350 responses	Sample varied by analysis; 1,328 in the initial psychometric analysis and 1,313 in current by-site analyses	Learner perceptions across high school experiences

Teacher surveys	141 responses	139 in the initial psychometric analysis; 141 available for subsequent descriptive analyses	Educator perceptions of learners' experiences
Learner focus groups	21 groups; 105 learners	Coded excerpts and scorable marker evidence	Learner descriptions across courses, time, and settings
Teacher interviews	37 interviews	Interview records	High school model and instructional context
High school materials	Varied by high school	Document or artifact	Intended structures, competency frameworks, and assessment systems
Learner outcomes	Varied by high school and measure	Learner, tested population, graduate cohort, or high school-level record	Descriptive evidence of learner progress and performance, primarily from 2024-25
Policy leader interviews and focus groups	17 conversations involving 45 participants: 10 site-focused conversations and 7 conversations with state agencies, nonprofits, and intermediaries	Conversation record; conversation-level presence for reported topic counts	Accounts of policy authority, local interpretation, implementation conditions, and support
Policy documents and contextual resources	Local document scan covering all ten sites; State of CBE policy map; broader landscape scan	Document, provision, or documented implementation arrangement	Evidence of formal policy conditions and their translation into local rules and systems

Classroom Observations

Participating high schools selected the classes and learning settings visited by researchers. Guidance asked high schools to identify learning experiences spanning grade levels, content areas, or important elements of their competency-based models. Settings could include core academic courses, applied learning, or other experiences central to the high school model. High schools were explicitly asked to maintain ordinary instruction rather than prepare a special lesson for the research visit.

A team of two FullScale researchers generally visited each high school for approximately a day and a half. Classroom observations lasted approximately 20 to 40 minutes and were noninteractive. Researchers observed learning as it occurred and did not question learners during the observation period.

Eight trained researchers participated in observation data collection. Two observers generally attended each classroom visit and independently completed a structured observation record. The finalized observation dataset included 49 classroom visits and 97 observer records. Forty-eight classroom visits produced two independent observer records; one classroom visit produced a single observer record.

Separate observer records allowed the study to examine both the evidence visible to each researcher from their position in the learning environment and the consistency and dependability of marker ratings across observers. *Appendix CE* describes observer training, the paired-observation design, marker observability, inter-rater analyses, observer effects, generalizability analyses, and the classroom observation protocol.

Student and Teacher Surveys

Participating high schools distributed electronic surveys through Qualtrics. Student surveys asked learners about experiences related to competency expectations, progression, assessment, feedback, applied learning, and the visibility of progress. Teacher surveys used parallel items to ask educators about the experiences available to learners within their classrooms and high schools.

Participation was voluntary. Respondents could skip individual questions or stop participating at any time. Partially completed surveys contributed responses to analyses for which the required items were available. Analytic sample sizes therefore varied across psychometric, cross-site, and by-site analyses.

The study received 1,350 student survey responses. The initial student psychometric analysis included 1,328 responses meeting the requirements for that analysis. Additional responses from one school were received after the initial psychometric analysis was completed. Current by-site analyses include 1,313 respondents meeting the inclusion requirements for those analyses. The differing counts reflect both the timing of data receipt and analysis-specific completeness and inclusion rules rather than attrition from one common analytic sample.

The study received 141 teacher survey responses. The teacher psychometric analysis included the 139 responses available when that analysis was completed; two additional teacher responses were received later. High school-level teacher samples varied substantially, so individual case profiles report teacher results descriptively or omit high school-level teacher findings when the available sample is too small for meaningful interpretation.

Appendix D documents student-survey development, scoring, item analysis, internal consistency, confirmatory factor analysis, validity evidence, and subgroup and by-site analyses. *Appendix E* provides corresponding teacher-survey methods and analyses. The complete student and teacher survey instruments appear in *Appendices H and I*, respectively.

Learner Focus Groups

Researchers conducted 21 learner focus groups with 105 learners across the ten participating high schools. High schools were asked to invite learners representing a mixture of grades, identities, interests, classrooms, and experiences with the local competency-based model. Recruitment was therefore purposive and facilitated by high school staff rather than random.

Focus groups generally included four to seven learners and lasted approximately 45 to 60 minutes. FullScale researchers facilitated the conversations using a semi-structured protocol connected to the hypothesized CBL Markers. The conversations invited learners to explain how they understood competencies, monitored their progress, responded to feedback, revised work, demonstrated mastery, and connected learning across experiences.

With permission, focus groups were audio-recorded and transcribed. One researcher coded the transcripts using the marker codebook and hypothesized breadth-of-experience continuum, and a second researcher reviewed the coding. Because coding was not conducted as two independent parallel coding processes, the focus-group analysis was not used to estimate inter-rater reliability. *Appendix F* describes participant recruitment, facilitation, transcription, coding, secondary review, analytic procedures, and the focus-group protocol in greater detail.

Focus groups provided a different window into learner experience than classroom observations. A brief observation could capture one learning episode; a focus-group participant could describe how an experience unfolded across several months, courses, educators, or settings. Focus-group evidence was therefore used to understand how learners interpreted and connected their experiences, not to estimate how frequently every learner within a high school had the same experience.

Teacher Interviews

Researchers conducted 37 teacher interviews across the participating high schools. High schools selected educators representing different grade levels, subjects, or roles within their competency-based models. Interviews generally lasted 30 to 45 minutes and addressed instructional practices, assessment, learner experience, and conditions surrounding implementation.

Teacher interviews were used primarily to understand local high school models and provide context for interpreting learner-experience evidence. They were not treated as a standardized measure for ranking marker presence across high schools.

High School Materials and Supplemental Learning Settings

Researchers reviewed materials including competency frameworks, rubrics, progress trackers, course documents, portfolios, and descriptions of exhibitions or presentations. These materials helped establish what participating high schools intended learners to experience and how competencies were represented within local systems.

Table A3 summarizes interview, focus-group, and supplemental evidence collected during the high school visits. Classroom-observation counts are reported at the aggregate level above and analyzed in detail in Appendix E.

Table A3. Qualitative and Supplemental Data Collection by High School

High School	Total Teacher Interviews	Total Learner Focus Groups	Total Focus-Group Learners
Empower[Ed]	3	3	13
Northern Cass High School	3	2	9
Crosstown High	3	2	10
STEAM Academy	4	2	8

Clark Street Community School	3	2	12
Harwood Union High School	3	2	8
Compass Community Collaborative School	4	2	9
Del Lago Academy	4	2	13
Da Vinci Communications	5	2	14
Lindsay High School	5	2	9
Total	37	21	105

Survey participation by high school is reported in *Appendices C and D* because the samples and appropriate forms of analysis differ by respondent group and site.

Policy

The policy analysis drew on 17 conversations involving 45 participants. Ten conversations provided direct accounts of local implementation, with one conversation focused on each participating site. Seven additional conversations involved state agency leaders, nonprofit organizations, and intermediaries working across the seven states represented in the study. These sources provided perspectives on both local implementation and the broader conditions surrounding it.

Two semi-structured protocols guided the conversations: a Policy Leader Interview protocol and a Policy Leader Focus Group protocol. Both addressed understandings of CBL, enabling and hindering conditions, variation across contexts, and opportunities for policy development. The interview protocol invited detailed accounts of a participant's context, including how schools navigated policy tensions and implementation tradeoffs. The focus-group protocol invited participants to compare experiences, examine shared patterns, and discuss variation across their contexts. The focus-group protocol allocated approximately 60 minutes for discussion.

Prompts addressed such topics as grading and credit, instructional time, assessment, accountability, waivers, postsecondary and workforce learning, and funding. These prompts provided a common basis for discussion while allowing participants to describe the policies and conditions most relevant to their experience.

The analysis also drew on a local document scan covering all ten sites, FullScale and KnowledgeWorks' [State of CBE](#) policy map, and the study's broader [landscape scan](#). Publicly available documents provided evidence of how policies were codified, interpreted, and communicated through state provisions, board policies, charter arrangements, administrative guidance, and school procedures. The conversations added accounts of how those conditions operated in practice.

Blueprinting Learner Outcomes With High Schools

The study did not begin with a uniform list of academic outcomes to collect across all ten high schools. Participating high schools differed in what they defined, prioritized, and tracked as evidence of learner progress. FullScale therefore used a collaborative blueprinting process to connect each high school's learning model with evidence that could responsibly describe learner performance.

Researchers first met with high school teams to develop logic models describing relationships among the local design, central practices, intended learner experiences, and expected outcomes. Logic models can make the relationship between an educational approach and its intended results explicit and provide a foundation for selecting measures meaningful to the program being studied (McLaughlin & Jordan, 1999). In this study, the logic models served as measurement blueprints and sources of contextual understanding; they were not treated as causal models tested by the study.

Researchers and high school teams then identified outcomes most proximate to each learning model. These conversations considered the evidence each high school used to understand whether learners were progressing, demonstrating competency, and preparing for graduation or postsecondary transitions. Depending on the high school, potentially relevant measures included competency progress, course completion, credits earned, GPA, benchmark and state assessment results, attendance, graduation, dual enrollment, career and technical education, work-based learning, and postsecondary enrollment.

The research team next developed an inventory of accessible data. Researchers and high school leaders discussed which data sources were available, which learners and time periods they represented, how measures were defined, what interpretations each source could reasonably support, and where missingness or inconsistent tracking limited interpretation. The team then identified required fields, reporting windows, export formats, and secure transfer procedures.

Outcome analyses primarily describe learner performance during the 2024–25 school year. Internal evidence was supplemented with public data when an appropriate reference point was available. Contextual comparisons were selected based on local relevance and could include another high

school, a district, a county, or a state. These groups were not statistically matched to the participating high schools, and comparisons are descriptive rather than causal.

At selected high schools, compatible learner-level records allowed researchers to examine relationships between local competency measures and conventional academic assessments. These analyses asked whether evidence generated within a local competency system was associated with other measures of academic performance. They did not test whether learners' survey-reported CBL experiences caused stronger outcomes.

Appendix F describes outcome selection, data preparation, measure definitions, denominators, contextual comparison choices, statistical procedures, and analytic limitations in greater detail.

Analyzing the Evidence

Researchers analyzed each source according to its structure and intended use before integrating findings across datasets.

Student- and teacher-survey analyses began with response distributions, item functioning, missingness, and marker-level descriptive statistics. Measurement analyses then examined whether groups of items provided sufficiently coherent evidence to support marker-level interpretation. The student survey also underwent confirmatory factor analysis and analyses of relationships with theoretically related constructs, including mastery- and performance-goal orientations, personalization, and project-based learning. Subsequent analyses used only marker scales supported by the measurement evidence. *Demonstration Location* items were interpreted individually because they did not form a sufficiently reliable composite scale.

Classroom-observation analysis examined how frequently each marker was scorable, the distribution of scorable evidence along the hypothesized breadth-of-experience continuum, agreement between paired observers, observer effects, and score dependability. "Insufficient evidence" indicated that an observation did not provide enough information to place the experience on the continuum. It was not interpreted as evidence that the marker was absent from the classroom or high school.

Learner focus-group transcripts were coded using the shared marker codebook. Analysts documented whether an excerpt provided codable evidence of a marker and, when appropriate, its location on the hypothesized breadth-of-experience continuum. Frequency summaries describe the available coded evidence rather than the percentage of all learners who experienced a marker.

For cross-source comparisons, survey markers were ordered by mean score, while observation and focus-group markers were ordered by the frequency with which sufficient evidence was available to assign a continuum level. These rankings summarize different properties of the evidence.

They were interpreted comparatively within each source, were not averaged across sources, and did not produce an overall high school or marker-quality score.

The policy analysis examined relationships between formal policy conditions and local implementation within and across the ten sites. It brought together participants' accounts and documentary evidence to describe available authority, how local actors translated that authority into operating arrangements, and where requirements or support structures shaped implementation. The analysis distinguished a policy provision's availability from evidence that a participating school actually used it.

The cross-site synthesis examined policy pathways, interactions among policy areas, implementation support, and relationships through which local experience informed broader policy learning. Illustrative cases showed how these conditions operated in particular contexts. The analysis did not assign schools policy-effectiveness scores or use policy evidence to rank the CBL markers or place learner experiences on the hypothesized breadth-of-experience continuum.

Where the policy analysis reported frequencies, the unit was the conversation. A count indicated whether an issue surfaced in a conversation, rather than the number of participants who raised it or the number of times it was discussed. Site-focused counts used ten conversations as the denominator, broader policy-context counts used seven, and combined counts used seventeen. These summaries describe coverage within the available evidence; they do not estimate the prevalence of a condition among CBL schools or policy leaders generally.

Specific analytic procedures, decision rules, reliability results, and supported uses of the measures appear in *Appendices C through H*.

Integrating Evidence Across Sources

The research team examined evidence within and across datasets to understand how learners perceived, described, and acted on CBL. Integration focused on where sources converged, where one source added explanatory context to another, and where findings diverged or remained incomplete.

The surveys provided the most consistent quantitative basis for comparisons across participating high schools. Classroom observations, learner focus groups, interviews, and high school materials were integrated alongside those results to show how learner experiences became visible in particular contexts and to illuminate dimensions that surveys could not capture directly.

The sources were not expected to produce identical results. A classroom observation might capture a learner revising one portion of a competency-aligned product without revealing the feedback received earlier or what happened afterward.

A learner could describe a project developing across several months and settings, but that account could not establish how commonly every learner experienced the same process. A rubric or competency framework could document an intended structure without demonstrating how learners understood or used it.

Triangulation therefore did not mean collapsing all evidence into a single score. It meant interpreting each source in relation to what it was designed and positioned to reveal. This approach is consistent with mixed-methods designs in which integration provides a fuller interpretation than any individual source can support independently (Fetters et al., 2013; Greene et al., 1989).

Participating high school teams reviewed draft site profiles to identify factual inaccuracies and clarify local terminology, data definitions, and model descriptions. The research team retained responsibility for analytic decisions and interpretation.

Policy evidence added context to the learner-experience findings by identifying the authority, operating rules, and support surrounding implementation. For example, a learner's description of reassessment could be interpreted alongside a school's grading procedures, while an account of community-based learning could be considered alongside rules governing how that evidence was assessed and recognized for credit. Connections were made where the sources supported them. Document evidence established formal provisions, participants described their interpretation and use, and learner-experience evidence showed particular experiences within those settings. Reading these sources together supported a fuller interpretation without establishing that a policy caused the experience.

Protecting Participants and Data

The study was reviewed under IRB Study 3354 and received expedited approval on April 10, 2026. It was classified as minimal risk.

Participation was voluntary. Learners and adult participants could decline participation, skip questions, or stop participating without consequences for academic standing, employment, or access to high school services. Families received study information and an opportunity to opt out of student surveys and noninteractive classroom observations. Learners received age-appropriate assent information, and learners who were 18 or older provided consent directly.

Participation in learner focus groups required parent or guardian permission for minors and learner assent. Learners who were 18 or older provided their own consent. Educators provided consent for surveys, interviews, and classroom observations as applicable. Focus-group and interview participants separately consented to audio recording.

Surveys were administered through Qualtrics. Internal learner data were transferred through encrypted or otherwise secure processes established with participating high schools and districts. Identifiers needed for data linkage and verification were removed or replaced with study codes before analysis. Focus-group and interview recordings were transferred to restricted storage, transcribed, de-identified, and deleted according to the approved protocol. Findings are reported in aggregate or through quotations identified only by a generic role and, when relevant, the participating high school.

Each participating high school received a \$10,000 institutional stipend, administered in two installments, to offset staff time and operational work required for data collection and review. Individual learners and educators were not compensated for participation.

The policy protocols emphasized voluntary participation and requested separate permission to record. The interview protocol specified that individuals would not be named without express consent and offered participants an opportunity to review and approve reporting that included their name or contributions. It also invited participants to identify comments that should remain confidential or unassociated with their context. The focus-group protocol asked participants to respect different perspectives, allow space for others, and keep the discussion confidential. This request established an expectation for the group; it could not guarantee confidentiality among participants.

Design Boundaries

The study design supports conclusions about which hypothesized CBL markers appeared across these ten intentionally varied high schools, how learners and educators perceived and described those experiences, and what variation in breadth was visible within the available evidence. It also supports descriptive accounts of the outcomes participating high schools prioritized and, at selected sites, correlational analyses between local competency measures and conventional academic measures.

Several features bound those interpretations. Participating high schools, observed learning settings, and learner focus-group participants were purposefully rather than randomly selected. Learner-experience data were collected primarily during a limited period in spring 2026, and classroom observations captured brief 20- to 40-minute episodes rather than repeated measures across an academic year. Focus groups provided depth about experiences across time and settings rather than population estimates, and survey participation varied across high schools.

Outcome analyses primarily relied on data from the 2024–25 school year and therefore did not represent outcomes measured concurrently with the spring 2026 learner-experience data. Outcome measures also varied because high schools maintained different data systems and prioritized different indicators.

Contextual reference groups were not matched, and observed relationships cannot establish causation. The study did not link survey-reported CBL experiences directly to academic outcomes due to sampling differences.

The policy evidence describes conditions surrounding the ten participating sites and perspectives from the leaders and organizations included in the conversations. It does not establish which policy arrangement is most effective or whether a particular provision is necessary or sufficient for CBL. Public documents describe formal authority and procedures but cannot, by themselves, establish how consistently those provisions operate in practice. Participants' accounts add implementation experience while remaining situated perspectives.

Policy conversation counts indicate whether a topic surfaced within the available discussion. A topic's absence does not demonstrate that the condition was absent or unimportant. The analysis also does not explain differences in marker scores, establish that policy produced broader learner experiences, or attribute academic outcomes to particular policy arrangements. Its contribution is to identify documented connections and questions that further research can examine.

These boundaries define where the evidence base can grow next. The markers, measures, and hypotheses developed through this study can now be tested and refined across additional high schools, broader learner samples, repeated observations, and longer periods of time. Longitudinal and comparative studies can further examine how competency-based learner experiences develop, how they vary within and across high schools, and how they relate to academic and broader learner outcomes. Technical limitations and the supported uses of individual measures are detailed in *Appendices C through H*.

Appendix B.

CBL Marker Development, Definitions, and Hypothesized CBL Continuum

The ten competency-based learning (CBL) markers provide a common lens for examining learner experiences across high schools with different competency structures, instructional models, grading systems, schedules, and terminology. Rather than defining CBL through any one visible structure, the markers focus attention on how learners encounter and use competencies through learning, assessment, and progression.

The markers began as theoretically grounded hypotheses informed by research, established definitions of competency-based education, field-developed models, and scholarship on learning and assessment. They were refined through a structured source crosswalk, independent researcher review, calibration, advisory feedback, and translation into survey items, observation criteria, and qualitative codes. This appendix describes that development process, defines the ten markers, and presents the shared hypothesized CBL continuum used to interpret classroom-observation and learner focus-group evidence. Technical information about the measures and analyses appears in *Appendices C–E*.

Developing the CBL Markers

Competency-based education (CBE) has been defined through multiple research traditions, policy agendas, and approaches to high school design. These sources share important commitments but do not always use the same language or clearly distinguish competency-based practices from related approaches such as personalized, student-centered, project-based, or whole-child learning. A systematic review of K–12 competency-based education similarly found variation in how CBE was defined and implemented, as well as identified a need for greater clarity about the practices that distinguish competency-based approaches from more conventional forms of schooling (Evans et al., 2021).

That variation shaped the study’s starting point. The research team reviewed sources with different purposes and units of analysis. Some defined CBE at the system level, while others described learner-centered environments, assessment practices, or the developmental conditions that support learning. Table B1 identifies the principal source areas represented in the crosswalk and explains how they contributed to the study’s learner-level hypotheses. The sources shown are representative rather than exhaustive.

Table B1. Representative Sources Informing Marker Development

Source Area	Representative Sources Documented in the Crosswalk	Contribution to Marker Development
Competency-based education definitions and design principles	Levine and Patrick (2019); Sturgis and Casey (2018); Education Reimagined (2020)	Established competencies as explicit and measurable expectations for learning. These sources informed hypotheses about progressing through demonstrated mastery, maintaining common expectations while varying time and support, making progress visible, and using assessment to guide continued learning.
Learner-centered and personalized learning	Big Picture Learning (n.d.); Education Reimagined (2020); LEAP Innovations (2024); Martin (2025); The Learning Accelerator (n.d.); Transcend (2025)	Focused attention on how learners experience educational structures. These sources informed hypotheses about organizing learning around competencies, recognizing competencies across contexts, documenting progress, drawing on evidence from different settings, and taking an active role in learning.
Assessment, application, and evidence of learning	Carnegie Foundation for the Advancement of Teaching (2025); Sturgis and Casey (2018); The Learning Accelerator (n.d.)	Informed hypotheses about how competency becomes visible through assessment, including transparent performance expectations, actionable feedback, additional opportunities to demonstrate learning, transfer across contexts, authentic application, and varied sources of evidence.
Whole-child and developmental perspectives	ASCD (2007); Germain et al. (2023); Stafford-Brizard (2016)	Situated competency development within the relational, developmental, and environmental conditions that shape learning. These sources helped the team maintain a learner-level focus while recognizing the broader conditions and supports upon which those experiences depend.

Note. Table B1 summarizes the principal bodies of work included in the development crosswalk rather than every individual source reviewed.

The crosswalk generated a broad set of potential learner experiences, instructional practices, and high school design features associated with CBL and related learner-centered approaches. Some frameworks described what learners experienced. Others described instructional practices or high school structures intended to support those experiences. Still others—including caring relationships, personalized pathways, learner choice, or community engagement—could be important within CBL environments but did not depend on learning being organized around competencies.

Researchers examined these ideas using two central questions. First, *did the idea describe something learners could encounter, understand, use, or demonstrate?* Second, *did the experience depend meaningfully on explicit competencies, or could it occur in much the same way within another learner-centered approach?* This distinction did not imply that relationships, personalization, choice, or other supportive conditions were unimportant. It helped identify learner experiences for which competencies provided the organizing logic.

Marker development was informed by established guidance emphasizing clear construct definitions, systematic examination of the relevant content domain, expert review, and iterative refinement before and during empirical testing (Boateng et al., 2018; Haynes et al., 1995). Three researchers independently reviewed the candidate ideas identified through the crosswalk for conceptual fit, level of specificity, and overlap with adjacent ideas. They compared their judgments and resolved differences through discussion. The marker set was then reviewed within the research team and with the study's advisory board before being translated into student- and teacher-survey items, classroom-observation criteria, and learner focus-group codes.

Through this process, closely related ideas were consolidated, while practices that could occur independently of competencies were treated as contextual conditions rather than distinct CBL markers. The process produced ten hypothesized learner-level markers of CBL, organized within two dimensions: **Learning Design** and **Assessment**.

Grounding the Markers in Competencies

An explicit competency structure provides the conceptual foundation for the ten markers, although participating high schools define and organize competencies in different ways. Some use cross-cutting competencies, graduate profiles, or durable skills intended to extend across courses and experiences. Others use course- or discipline-specific competencies. In several high schools, academic standards function as competencies because they organize expectations, evidence, assessment, progression, or demonstrations of proficiency across contexts and domains. Competencies may therefore encompass a combination of academic knowledge, transferable skills, and dispositions.

The high school model may define competencies explicitly even when that connection is not visible to a learner in every learning context. The presence of a standard or competency on an assignment also does not, by itself, determine the learner experience. The markers examine what happens around that expectation: whether learners connect competencies to their work, understand and document progress, use feedback to improve evidence, apply learning through meaningful demonstrations, and experience consequences when proficiency is demonstrated.

The two dimensions offer complementary views of those experiences:

- **Learning Design** examines how competencies organize learners' understanding of expectations, development, progress, and movement through learning.
- **Assessment** examines how evidence of competency is generated, strengthened, demonstrated, interpreted, and evaluated.

These dimensions are related but are not stages in a causal sequence. Together, they provide two ways of examining how competencies become consequential in learners' experiences.

Defining the Ten Markers

Table B2 defines the learner-level experience represented by each marker. The markers emerged from a crosswalk of research and field-based descriptions of competency-based and learner-centered education. Across these sources, the research team identified shared ideas about explicit competencies, progress based on demonstrated mastery, transparent expectations, varied sources of evidence, meaningful application, and assessment that supports continued learning. Frequently referenced sources included definitions and quality principles developed by the Aurora Institute, along with resources from Education Reimagined, The Learning Accelerator, LEAP Innovations, and the Carnegie Foundation for the Advancement of Teaching (Carnegie Foundation for the Advancement of Teaching, 2025; Education Reimagined, 2020; LEAP Innovations, 2024; Levine & Patrick, 2019; Sturgis & Casey, 2018; The Learning Accelerator, n.d.).

The research team synthesized and refined these ideas into ten markers focused specifically on what learners may experience within competency-based environments. Because the source materials differed in purpose, terminology, and level of analysis, no single source maps directly onto the final set. The definitions in Table B2 represent hypotheses developed for this study and subsequently translated into survey items, observation criteria, and focus-group codes.

Table B2. Definitions of the Ten Hypothesized CBL Markers

Dimension	Marker	Definition
Learning Design	 Mastery-Based Progression	The extent to which demonstrated proficiency influences when learners move forward and where they continue working.
Learning Design	 Clear Articulation	The extent to which learners organize and make meaning of school and lived experiences around cross-cutting competencies.
Learning Design	 Developmental Progression	The extent to which learners recognize and engage with increasing complexity and depth within a competency over time.
Learning Design	 Connection to Learning Targets	The extent to which learners connect content- or skill-specific learning targets to broader competency development.
Learning Design	 Transparency	The extent to which learners can access and use records of competency evidence to understand their progress and what remains.
Learning Design	 Universal Application of Expectations	The extent to which learners recognize and use common competency-based performance criteria across different work, educators, or contexts.
Assessment	 Iterative Purpose	The extent to which learners seek and use adult and peer feedback, revise through multiple drafts or attempts, and request targeted feedback to strengthen evidence toward mastery.
Assessment	 Rigorous Application of Learning	The extent to which learners produce authentic, contextually meaningful work that serves as evidence of mastery aligned with competency criteria.
Assessment	 Demonstration Location	The extent to which learners identify, collect, and submit evidence of competency from school and lived experiences.
Assessment	 Clear Calibration	The extent to which competency evidence is interpreted using shared criteria intended to support accurate and consistent judgments across assessors.

Developing a Hypothesized CBL Continuum

The study was designed to examine more than whether evidence related to a marker appeared. The research team also wanted language for describing how learner experiences varied when a marker was visible.

Consider two experiences related to the same competency. In one, an educator points learners to a competency on an assignment and explains its relationship to the current task. In another, a learner independently draws evidence from several courses, a workplace experience, and a community activity and explains how those experiences demonstrate the same competency. Both provide evidence relevant to CBL, but the role of the competency, learner ownership, and scope of the experience differ.

The research team therefore developed a four-level hypothesized CBL continuum defining the breadth or range of learner experiences: *Pre-Emergent*, *Baseline*, *Maturing*, and *Advanced*. The continuum was designed to describe variation within a particular piece of evidence rather than to classify a learner, classroom, or high school as more or less competency-based.

The continuum varies along three dimensions:

- **Role of competencies:** whether the competency connection is implicit, explicitly referenced, applied with support, or independently integrated into learning.
- **Learner ownership:** whether the experience is primarily adult-directed, adult-supported, shared between learner and adult, or learner-directed.
- **Scope:** whether the experience is bounded to a single task, situated within one course or experience, connected across multiple experiences, or integrated across settings and over time.

The three dimensions informed a holistic, best-fitting judgment; they were not scored or averaged separately. Evidence was coded *Pre-Emergent* only when researchers had enough information to identify an experience relevant to the marker but found limited evidence of competency connection, learner ownership, or scope. When a classroom-observation record or focus-group excerpt did not contain enough information to interpret the marker, it was classified as *insufficient evidence* and was not placed on the continuum.

Table B3. Hypothesized CBL Continuum

Dimension	Pre-Emergent	Baseline	Maturing	Advanced
Role of competencies	A relevant learning experience is visible, but the competency connection is implicit within the evidence.	Competencies are explicitly referenced.	Competencies are applied with support.	Competencies are independently integrated into learning.
Learner ownership	Adult-directed	Adult-supported	Shared ownership	Learner-directed
Scope	Single task or moment	Single course or bounded experience	Multiple experiences	Multiple school and lived settings over time
Overall interpretation	The learner participates in an experience relevant to the marker, but the competency connection is limited within the available evidence.	The learner encounters an explicit competency and related structures, generally within an adult-supported context.	The learner uses competencies and evidence across multiple experiences with structured support or shared decision-making.	The learner independently uses competencies and evidence to organize, interpret, or direct learning across settings and over time.

The continuum is a hypothesis rather than a validated developmental scale. Its levels do not represent stages through which every learner is expected to progress, nor does an *Advanced* rating indicate that the experience is always the most appropriate form of instruction. Direct teaching, common timelines, educator guidance, scaffolding, and structured practice can all play important roles within CBL. The continuum provides an initial language for describing variation in the breadth of learner experience that can be tested and refined through future research.

Using the Markers and Continuum in This Study

The ten markers informed the student and teacher surveys, classroom-observation protocol, and learner focus-group codebook, although each source operationalized them differently. Surveys measured respondents' perceptions of experiences associated with the markers but were not scored using the CBL continuum. The continuum was applied only to classroom-observation records and coded learner focus-group excerpts when enough evidence was available to support interpretation. Detailed measurement and analytic procedures appear in *Appendices C–E*.

The continuum was applied to individual pieces of evidence, not to learners, classrooms, or high schools as a whole. A classroom observation or focus-group excerpt placed at one level does not establish that the learner consistently experienced that level across courses, settings, or time. Likewise, insufficient evidence in a brief classroom observation does not establish that the marker was absent from the classroom or high school.

The distinction between a conceptual marker and its measurement is also important. A marker can remain conceptually meaningful even when a particular instrument does not measure it adequately. In this study, psychometric analyses supported reliable student-survey composite measures for nine marker dimensions. The *Demonstration Location* items did not form a sufficiently reliable composite and were therefore interpreted individually rather than combined into a scale. This finding identifies a limitation of the current survey measure rather than determining whether *Demonstration Location* belongs within the conceptual marker set. Detailed student-survey findings appear in *Appendix C*.

Across the ten participating high schools, *Mastery-Based Progression*, *Iterative Purpose*, and *Rigorous Application of Learning* had the strongest, recurring cross-site evidence across the study's data sources. Their prominence emerged empirically; the three markers were not designated in advance as conceptually more fundamental than the other seven, nor were they necessarily the most prominent markers within every participating high school. All ten remain part of the study's hypothesis about the ways learners may experience CBL.

Interpreting the Marker Set

The markers provide a shared descriptive lens for examining how learners encounter and use competencies. They do not encompass every instructional practice, relationship, or high school condition that may contribute to a strong learning environment. Likewise, the breadth continuum describes what is visible within a particular piece of evidence; it does not rate learner ability, educator effectiveness, or the overall quality of a high school.

The four continuum levels remain hypothesized descriptive anchors rather than validated developmental stages. Learners may encounter different levels of breadth across markers, courses, settings, and points in time. Greater learner ownership is also not synonymous with better instruction in every context. Adult guidance, explicit instruction, scaffolding, structured practice, and feedback may all be appropriate within competency-based learning.

Current evidence therefore supports using the marker set and hypothesized continuum as descriptive and research tools for examining learner experience. Consistent with contemporary measurement theory, the defensibility of an interpretation depends on the evidence supporting that particular use rather than residing permanently in the measure itself (Kane, 2013). At this stage, the marker set does not support universal thresholds, high-stakes classifications, or high school-level quality judgments.

The contribution of this appendix is to make the conceptual hypotheses, marker definitions, and shared breadth continuum explicit enough for continued testing. Future research can examine how the markers appear across additional high schools and learner populations, how experiences change over time and across settings, and whether the hypothesized continuum remains useful as the evidence base grows. Making these assumptions and boundaries visible creates a foundation the field can continue to test, challenge, and refine.

Appendix C.

Psychometric Analysis for Observation Protocol

Purpose and Research Question

The CBL Observation Protocol was developed to support systematic documentation of learner behaviors associated with the ten hypothesized competency-based learning (CBL) markers. The protocol was field tested across the ten participating high schools and was designed to capture evidence visible during brief classroom observations rather than to provide a comprehensive measure of an educator's practice, a classroom's implementation, or a high school's overall competency-based model.

This appendix examines the measurement properties of the observation protocol and addresses the following research question and hypotheses:

- RQ1: What evidence supports valid interpretations and inferences drawn from the classroom observation protocol about observable markers of competency-based learning?
- H1: Examination of score variance and error variance indicates that the observation protocol captures meaningful differences in competency-based learning across classrooms and schools.
- H2: Results from generalizability analyses indicate that variance attributable to potential sources of measurement error (e.g., observer) is sufficiently accounted for in interpretation of ratings.

The analyses examined four related aspects of the observation evidence: the extent to which individual markers were observable during classroom visits, agreement between paired observers, sources of variation in ratings, and the dependability of ratings under the implemented observation design.

Methods

The CBL Observation Protocol was developed to understand the classroom implementation of competency-based structures through observable learner behaviors. The protocol operationalizes theoretically derived indicators of classroom environments expected to foster mastery-oriented, autonomous, and self-directed learning among adolescents. In contrast to most classroom observation systems, which focus primarily on teacher practices (e.g., CLASS), the protocol conceptualizes learners and their learning as the primary unit of analysis.

It documents how students interpret, use, and respond to competency-based structures through observable cognitive and behavioral attributes, extending existing observation approaches beyond engagement or instructional breadth to capture the implementation of competency-based learning.

This learner-centered conceptualization is consistent with classroom observation approaches that infer characteristics of learning environments through observable learner behaviors. For example, Cognia's Effective Learning Environments Observation Tool (Cognia, 2023) characterizes learner-centered environments by documenting how learners engage with instruction, interact with peers, demonstrate agency, and respond to classroom conditions. The CBL Observation Protocol extends this learner-centered perspective by focusing specifically on the instructional design features of competency-based education. The protocol examines how learners interpret and strategically engage with these features as evidence of the breadth of learning experiences taking place in CBL environments.

Learner Attributes

When developing the protocol, each CBL instructional design marker was operationalized as a developmental continuum describing increasingly sophisticated learner thinking and behavior expected as school-level competency-based structures become more coherent and classroom-level instruction increasingly supports mastery-oriented, autonomous, and self-directed learning. For each marker, the protocol distinguishes cognitive attributes (what learners understand, interpret, or believe) from behavioral attributes (what learners do), enabling observers to document both learners' meaning-making when verbalized and their strategic actions during instruction.

The cognitive dimension captures how learners interpret and make meaning of competency-based structures, such as understanding why cross-cutting competencies matter, explaining how competence develops through increasingly sophisticated levels of performance, interpreting progress as contingent on demonstrated mastery rather than seat time, or viewing feedback as information for improvement. Learners' thinking cannot always be observed directly. Therefore, cognitive attributes are to be inferred from learners' verbalizations, explanations, reflections, and interactions with peers and educators.

The behavioral dimension captures how learners strategically act on those understandings, such as organizing learning experiences around competencies, selecting increasingly challenging learning opportunities, using competency progressions to guide future learning, maintaining portfolios or progress trackers, seeking formative feedback before completing a summative assessment, or documenting evidence of learning from school and community contexts. Behavioral attributes are to be documented through observable actions, decision-making, and artifacts produced during instruction.

Developmental Continuum

Within each domain (architecture of learning and architecture of assessment), markers were rated along a four-level developmental continuum representing increasingly sophisticated implementations of competency-based learning: 1. Pre-Emergent CBL, 2. Baseline CBL, 3. Maturing CBL, and 4. Advanced CBL. The rating scale reflects the extent to which the structural and instructional conditions of the classroom make it reasonable to expect learners to demonstrate increasingly sophisticated patterns of cognition and behavior associated with competency-based learning.

At the Pre-Emergent level, learners are expected to organize their learning primarily around tasks, grades, teacher direction, or seat time, reflecting little evidence of competency-based structures. Baseline CBL reflects emerging evidence that learners recognize and begin to use competency-based structures, although their understanding and actions remain largely dependent on adult guidance and instructional supports. Maturing CBL reflects learners who increasingly interpret and strategically use competency-based structures to guide their learning, though independence and consistency continue to develop. Advanced CBL represents classrooms in which learners routinely and independently use competency-based structures to regulate, monitor, and direct their learning across academic and lived experiences, demonstrating mastery-oriented, autonomous, and self-directed learning patterns.

Observation Procedures

Observers were assigned to classrooms in fixed two-person teams. Across the 10 schools studied, 49 classrooms were observed by 8 observers who were organized into 6 observer pairs. Learner behaviors in each classroom were observed by one observer pair.

Analysis

Descriptive analyses examined the observability of each marker. The team employed summary statistics and distribution of scorable ratings across the CBL continuum, and inter-rater agreement using Cohen's kappa. Next, Generalizability Theory (Brennan, 2001) was applied to examine the dependability of the *observation procedures* as they were implemented in this project. It helps to understand the contributions of known sources of error variance such as observers to variance in breadth of learner experience CBL ratings assigned to learner behaviors in classrooms.

Inspection of the assignment structure indicated several observers participated in observations across multiple schools and classrooms. For example, observer RL participated in observations within three different observer pairs and across multiple schools, while observer MB also participated in different observer pairs. This overlap created a connected observer structure, allowing us to proceed with a partially-crossed mixed-effects generalizability study.

The univariate generalizability analyses estimated variance attributable to:

- classroom (object of measurement),
- individual observer (source of error variance), and
- residual error (including classroom-by-observer interaction and other unexplained variance).

The single-facet, mixed-effects model for each competency-based learning marker was specified as:

$$\text{CBL Rating}_{ij} = \mu + c_i + r_j + e_{ij},$$

where c_i represents classroom effects, r_j represents the observer facet, and the residual term captures classroom-by-observer interaction and other unexplained sources of variation. All "insufficient evidence" ratings were excluded from the model as they do not provide relevant information about the measurement scale (i.e., pre-emergent cbl, baseline cbl, etc.)

A limitation of the univariate approach is that it does not allow covariation between markers to be modeled, as several are likely moderately correlated in practice. Future studies should consider a multivariate generalizability approach given the conceptual relationship among markers. However, given the partially crossed observation design, unequal marker observability, and the substantial amount of structural missingness resulting from context-dependent opportunities to observe specific learner behaviors, the univariate approach provided the most appropriate and interpretable analytic framework for the current study.

Finally, since observer pairs were assigned to specific schools, an additional model including a random school effect was estimated for one of the markers to determine whether school-level differences explained meaningful additional variation in ratings. Model comparisons indicated that including a school-level random effect did not improve model fit. Thus, a school effect was not included in subsequent models.

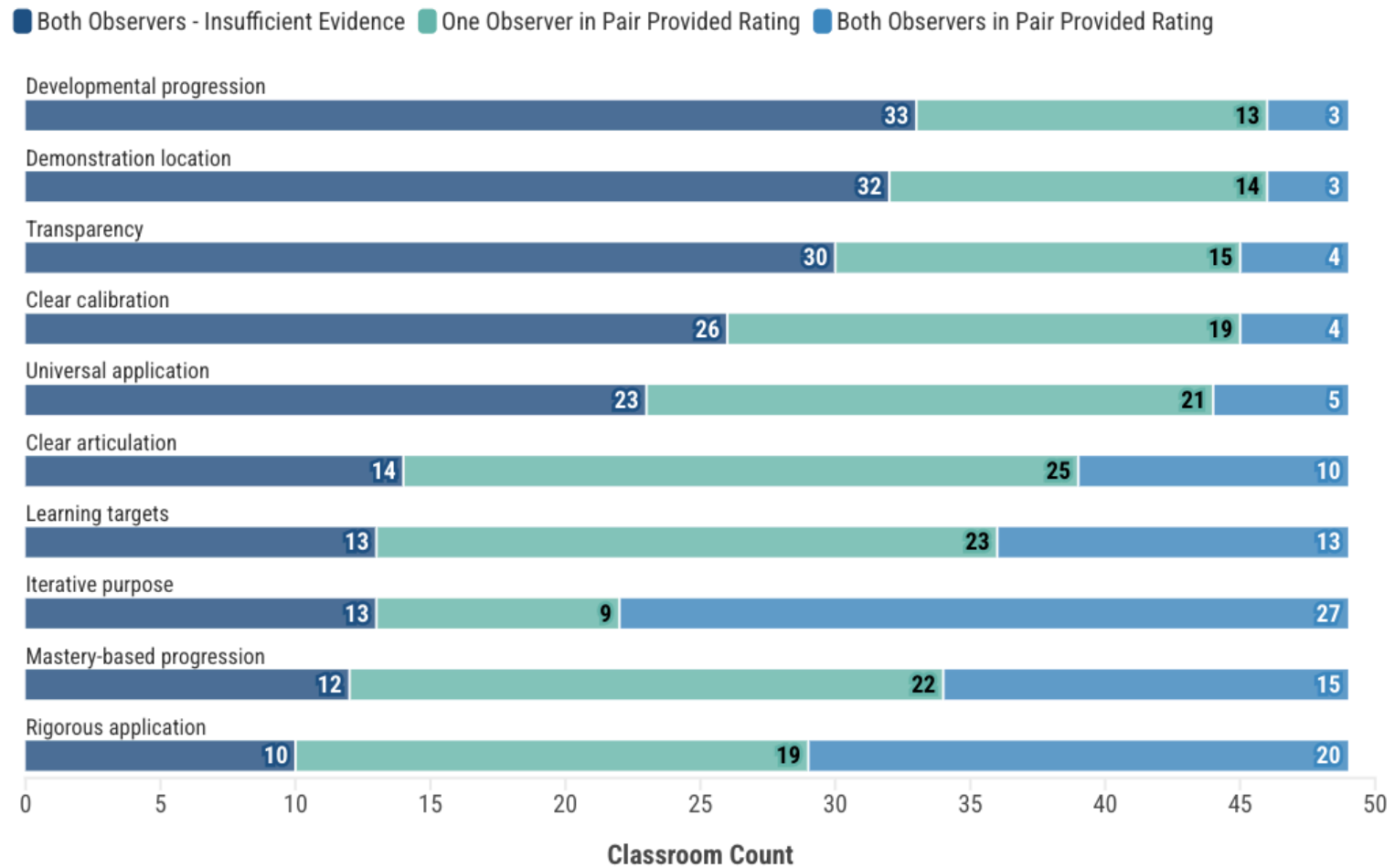
Findings

Across the ten markers, the proportion of observable behaviors varied considerably, further prompting caution in generalizing to overall implementation based on a single observation.

- Markers such as *Iterative Purpose* and *Rigorous Application of Learning* were frequently observed and rated, with approximately half of classrooms receiving ratings from both observers in the pair (Figure 1).
- Markers such as *Developmental Progression*, *Any Demonstration Location*, *Transparency*, and *Calibration* were **rarely** observed by both raters during a single classroom observation.

The low frequency of observation of several markers suggests that they may require alternative evidence sources (e.g., advisory observations, learner interviews, or artifact review) or multiple observation occasions to be measured reliably.

Figure 1. Observability of CBL Markers Across Classrooms



Descriptive statistics for scorable ratings are presented in Table 1. Median ratings were in the “baseline CBL” category for most markers. Six markers exhibited the full scale range (pre-emergent to advanced CBL), indicating that observers used the complete scale when sufficient evidence was available. However, the range was significantly restricted for *Developmental Progression*, where the highest category used was “baseline CBL.” Standard deviations ranged from 0.48 to 1.07, indicating modest variability in ratings across scorable observations.

Table 1. Descriptive Summary of Scorable Ratings (Classroom-Rater Observations) by CBL Marker

CBL Marker	n	Mean	SD	Median	Min	Max
Clear Articulation	45	1.8	0.79	2	1	4
Developmental Progression	19	1.32	0.48	1	1	2
Learning Targets	49	1.65	0.75	2	1	4
Transparency	23	1.96	0.64	2	1	3
Universal Application	31	1.71	0.74	2	1	4
Mastery-Based Progression	52	2.02	0.87	2	1	4
Demonstration Location	20	2.15	0.75	2	1	3
Clear Calibration	27	1.96	0.71	2	1	3
Iterative Purpose	63	2.48	0.95	2	1	4
Rigorous Application	59	2.14	1.07	2	1	4

* 1 = Pre-emergent CBL, 2 = Baseline CBL, 3 = Maturing CBL, and 4 = Advanced CBL

Shown in Figure 2 below, is the distribution of ratings across the four scale categories for each marker. Ratings spanned the full range of the scale for most markers, although the distribution differed across markers.

- *Rigorous Application* and *Iterative Purpose* exhibited the broadest distribution of ratings, including the greatest number of observations classified as advanced CBL.
- All of the remaining markers were concentrated in the pre-emergent and baseline categories, with few ratings in the upper scale categories

Across markers, relatively few observations were classified as “advanced CBL,” **suggesting either that learner behaviors associated with these markers were infrequently observed during classroom visits or that the criteria for the highest scale category were applied conservatively by observers.**

One unexpected finding was the relatively high count of pre-emergent and baseline CBL ratings for *Mastery-Based Progression*, where students progress based on demonstration of mastery rather than seat time. Schools were purposefully selected for this study because they are widely recognized for competency-based learning, where mastery-based progression tends to be recognized as a defining feature.

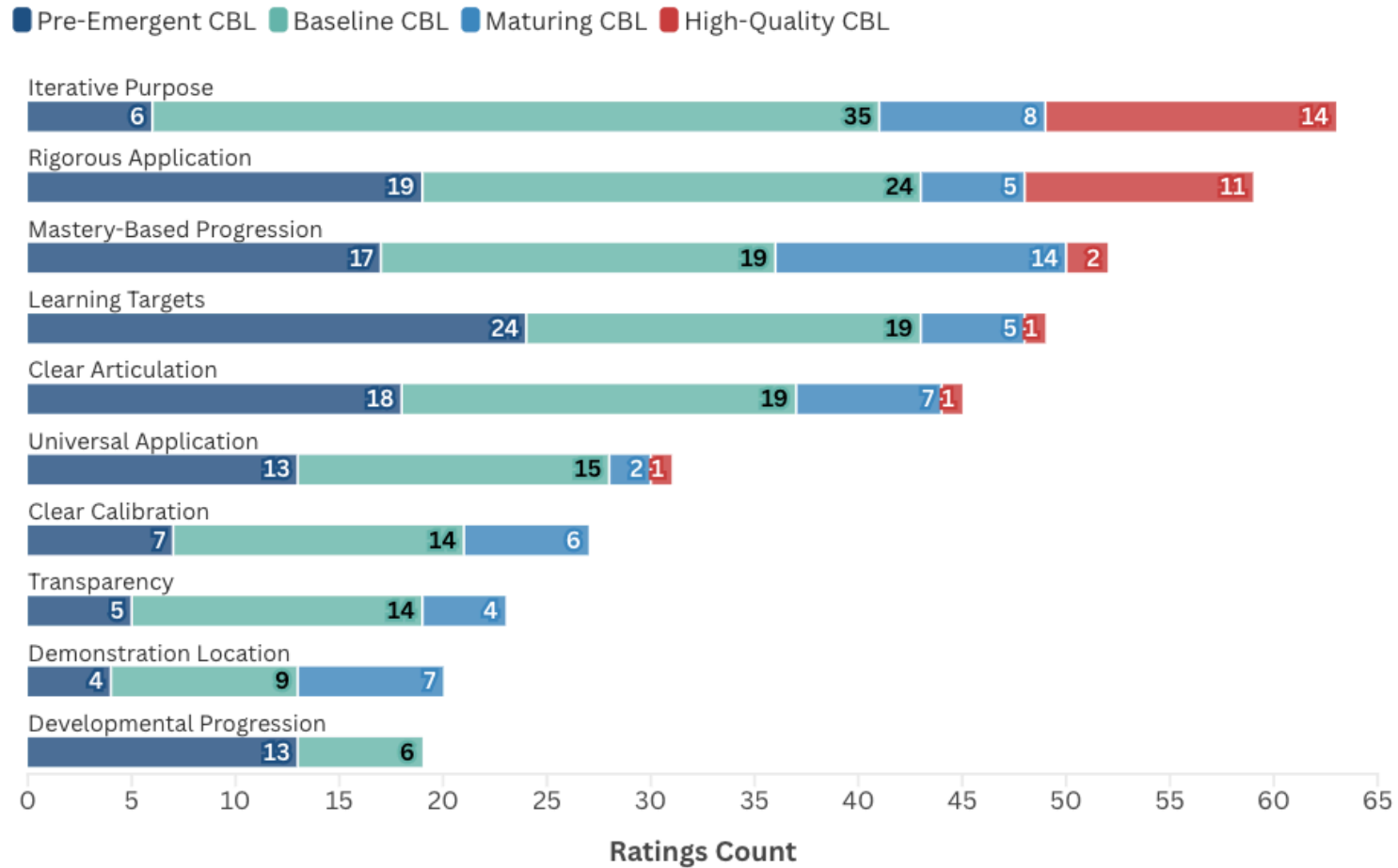
This discrepancy highlights the distinction between the presence of schoolwide competency-based structures and the extent to which mastery-based progression is directly observable within individual ~30-minute classroom observations.

Both the landscape scan and school intake survey indicated that participating schools had adopted many of the organizational structures commonly associated with competency-based learning. Schools consistently reported having competency frameworks, transparent learning targets, rigorous expectations for all learners, assessment practices intended to support learning, and learner-centered instructional approaches. Publicly available materials similarly emphasized competency frameworks, grading systems, learner profiles, and assessment philosophies.

Yet these organizational commitments seem to not have translated into consistently observable evidence of mastery-based progression during classroom observations. Further, even though most schools reported that mastery-based progression was part of their instructional model, observations frequently yielded Pre-Emergent or Baseline ratings for this marker. This finding, coupled with the low observability of five other markers, suggests most schools are still in the early stages of implementing core competency-based structures.

The purposeful selection of *exemplary* schools makes these findings particularly informative. If mastery-based progression is not consistently observable within classrooms in schools recognized as leaders in competency-based learning, then its presence should not be assumed in studies of broader populations of schools. More fundamentally, the findings suggest that school-level designation as “competency-based” and classroom-level enactment competency-based learning represent related but distinct constructs that require different forms of evidence to measure.

Figure 2. Distribution of Scorable Observation Ratings Across CBL Continuum



Inter-rater Agreement

Inter-rater agreement analyses were conducted to examine the extent to which paired observers made similar rating decisions. Cohen's kappa (Cohen, 1960) was estimated for markers with paired ratings from more than two classrooms. Agreement was calculated both including and excluding "Insufficient Evidence" ratings to examine the influence of observers' use of this designation. Kappa coefficients ranged from 0.03 to 0.37 when "Insufficient Evidence" ratings were included and from 0.00 to 0.44 when they were excluded, indicating **poor to moderate agreement overall**. Developmental Progression was excluded because too few paired scorable ratings were available to estimate a stable coefficient.

The modest improvement in agreement after excluding "Insufficient Evidence" ratings suggests that paired observers did not consistently apply this designation. This inconsistency may reflect differences in what observers were able to see within the classroom, such as variation in observer positioning or access to instructional interactions, rather than disagreement about the breadth of observed behaviors.

Mixed-Effects Model

Next, variance components for ratings, observers, and residual error were estimated for each CBL marker (Table 2). Among the six markers that yielded stable models, the relative contribution of classroom, observer, and residual variance differed substantially, indicating that the dominant sources of measurement variation varied by marker.

- *Mastery-Based Progression* was characterized primarily by classroom-level variance (59%), suggesting meaningful differences among classrooms with relatively little observer-related variation in how they applied the rubric.
- *Learning Targets* (40%) and *Rigorous Application* (48%) were associated with substantial observer variance, indicating that ratings were more sensitive to systematic differences among observers in how they applied the rubric for these markers.
- *Any Demonstration Location* and *Iterative Purpose* were characterized by relatively large residual variance (85% and 55%, respectively), suggesting that a considerable proportion of variability remained unexplained by classroom or observer effects.
- *Clear Articulation* was associated with substantial observer (37%) and residual (45%) variance, indicating relatively inconsistent judgments across observers and unaccounted for error variance.

The analyses also demonstrated that the psychometric estimability of the markers varied under the current observation design.

- Six markers (*Clear Articulation, Learning Targets, Mastery-Based Progression, Clear Calibration, Iterative Purpose, and Rigorous Application*) yielded stable variance components.
- *Developmental Progression* did not provide sufficient replicated observations to support reliable estimation.
- *Transparency, Universal Application, and Clear Calibration* produced models that failed to converge and were nearly unidentifiable, indicating that the available data were insufficient to support stable variance estimation.

Altogether, these findings suggest that the current observation design supports more dependable measurement for some competency-based learning markers than others. They also indicate that future iterations of the protocol may benefit from a more focused observation design that prioritizes markers demonstrating adequate observability and psychometric stability while continuing to refine the behavioral indicators, evidence rules, and scoring guidance for markers that could not yet be measured dependably.

Table 2. Decomposition of Variance in Classroom-Rater Observation Ratings and % Contribution

CBL Marker	Classroom-Rater Observations, n	Classroom Variance (%)	Observer Variance (%)	Residual Variance (%)
Clear articulation	45	0.109 (17.7%)	0.231 (37.4%)	0.277 (44.9%)
Developmental progression	19	*		
Learning targets	49	0.266 (41.7%)	0.255 (40%)	0.177 (18.4%)
Transparency	23	†		
Universal application	31	†		
Mastery-based progression	52	0.411 (59.1%)	0.068 (9.7%)	0.217 (31.2%)
Any demonstration location	20	0.006 (1.1%)	0.078 (13.6%)	.493 (85.3%)
Clear calibration	27	†		
Iterative purpose	63	0.131 (14.8%)	0.270 (30.5%)	0.484 (54.7%)
Rigorous application	59	0.362 (33.5%)	0.520 (48.1%)	0.199 (18.4%)

*Insufficient replicated observations to estimate variance components reliably

† Model did not converge and is nearly unidentifiable; variance estimates are unstable and not reported.

Generalizability and Dependability Studies

A generalizability study was conducted for the six markers that yielded stable variance component estimates to evaluate the dependability of classroom ratings under the current two-observer design (Table 5). Generalizability (G) coefficients (E_p^2) varied substantially across markers, ranging from 0.272 for *Rigorous Application* to 0.798 for *Mastery-Based Progression*, indicating that the protocol provided different levels of measurement precision depending on the marker.

- Only *Mastery-Based Progression* exceeded the commonly cited benchmark of 0.70 for acceptable reliability in relative decisions (Shavelson & Webb, 1991).
- *Learning Targets* (0.58) and *Rigorous Application* (.50) demonstrated moderate dependability.
- *Clear Articulation* (0.30), *Iterative Purpose* (0.26), and *Any Demonstration Location* (0.02) exhibited low dependability under the current observation design.

In practical terms, the current two-observer design does not provide sufficient reliability to support ranking classrooms on *Clear Articulation*, *Iterative Purpose*, and *Any Demonstration Location*, in particular, suggesting that additional observers, repeated observations, or refinements to the observation protocol would be needed to improve score dependability.

Table 5. Decomposition of Variance Components and Dependability Coefficient (E_p^2)

CBL Marker	n	Classroom Variance, n (%)	Observer Variance, n (%)	Residual Variance, n (%)	Reliability, E_p^2
Clear articulation	45	0.109 (17.7%)	0.231 (37.4%)	0.277 (44.9%)	0.301
Developmental progression	19	*			
Learning targets	49	0.266 (41.7%)	0.255 (40%)	0.177 (18.4%)	0.588
Transparency	23	†			
Universal application	31	†			
Mastery-based progression	52	0.411 (59.1%)	0.068 (9.7%)	0.217 (31.2%)	0.798
Any demonstration location	20	0.006 (1.1%)	0.078 (13.6%)	.493 (85.3%)	0.022
Clear calibration	27	†			
Iterative purpose	63	0.131 (14.8%)	0.270 (30.5%)	0.484 (54.7%)	0.257

Rigorous application	59	0.362 (33.5%)	0.520 (48.1%)	0.199 (18.4%)	0.501
----------------------	----	---------------	---------------	---------------	-------

*Insufficient replicated observations to estimate variance components reliably

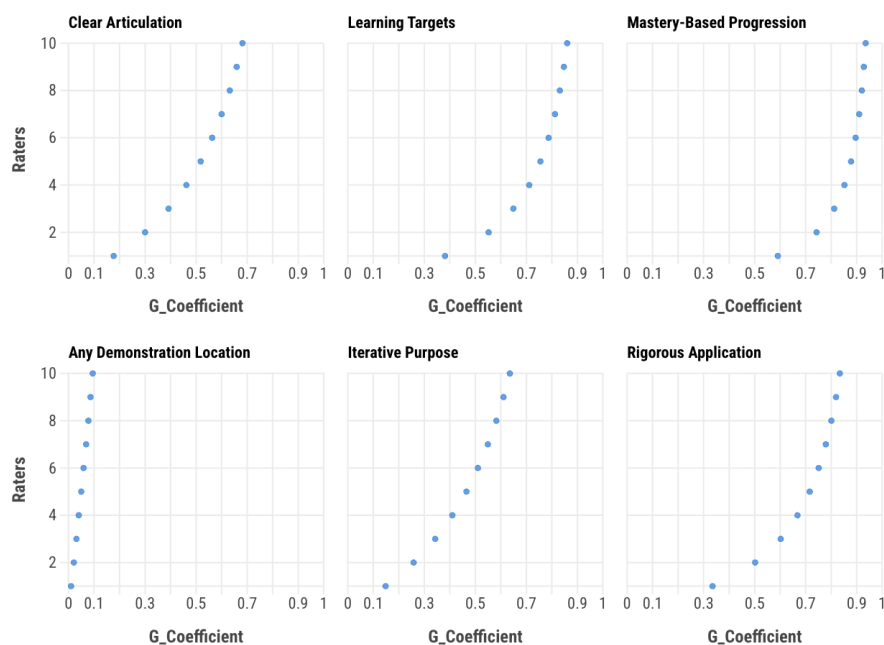
† Model did not converge and is nearly unidentifiable; variance estimates are unstable and not reported.

The D-study results are shown in Figure 5. As expected, increasing the number of observers improved the generalizability coefficient for all six markers; however, the rate of improvement varied considerably across markers.

- *Mastery-Based Progression* achieved acceptable relative dependability ($G \geq .70$) with a single observer and exceeded .90 with multiple observers, reflecting its large classroom-level variance and relatively small observer-related error.
- *Learning Targets* and *Rigorous Application* reached acceptable dependability with approximately three observers, whereas *Clear Articulation*, *Iterative Purpose*, and *Any Demonstration Location* required substantially greater observer replication to approach conventional reliability benchmarks. Even with ten observers to a single observation, gains for these markers remained gradual, indicating that observer replication alone is insufficient to overcome the large residual and observer-related sources of measurement error.

These findings suggest that improving the dependability of the observation of these markers will require changes to the measurement design beyond simply increasing the number of observers, such as refining behavioral indicators and scoring guidance, increasing opportunities to observe the target practices, or supplementing classroom observations with complementary sources of evidence.

Figure 5. Dependability study results by CBL Marker



Discussion

The purpose of these analyses was to understand the psychometrics properties of the CBL Markers Observation Protocol and describe patterns across observations. The CBL Markers Observation Protocol was developed through iterative refinement and field testing across 10 diverse competency-based learning environments.

Preliminary psychometric analyses demonstrated **meaningful variation in marker observability, inter-rater agreement, and score dependability**, indicating that additional validation is needed before observation ratings can be interpreted as stable measures of competency-based learning or used for high-stakes decision-making.

One unexpected finding was the relatively high count of “pre-emergent” and “baseline” ratings for Mastery-Based Progression despite the purposeful selection of schools recognized for competency-based learning. Variance decomposition indicated that ratings for this marker were driven primarily by classroom-level variation (61.4%) with relatively little observer-related variance (7.5%), which suggests the pattern reflects meaningful differences in classroom practice rather than inconsistent observer judgments. **These findings challenge the assumption that competency-based learning is enacted uniformly across classrooms within competency-based schools.** Instead, they highlight an important distinction between system-level implementation and classroom-level enactment: schools may establish structures that support competency-based learning, yet the extent to which mastery-based progression is observable may vary across classrooms, lessons, and instructional contexts.

More broadly, these findings suggest that **competency-based learning is best understood as a hierarchical construct**. School-level policies and structures create conditions for competency-based learning, but classroom observations capture its instructional enactment within individual lessons. This has implications for the evidence and measurement framework required to fully understand the extent to which schools are progressing toward competency-based learning.

The findings also have implications for the design of future classroom observations. Under the observation procedures used in this study, six markers—*Clear Articulation, Learning Targets, Mastery-Based Progression, Any Demonstration Location, Iterative Purpose, and Rigorous Application of Learning*—were consistently observable and yielded sufficiently stable measurement properties to support preliminary psychometric analyses. Future iterations of the protocol may benefit from a more focused design that prioritizes these markers during observations while continuing to **refine evidence rules and observation procedures for less frequently observed markers**.

Concentrating observations on a smaller set of consistently observable behaviors may allow observers to collect richer evidence, make more consistent judgments, and improve the dependability of observation data while additional validation work is conducted.

Preliminary Recommendations on Observation Protocol Use

The psychometric findings inform how classroom observation data should be interpreted and used throughout the remainder of this study.

Analytic Guidance for Full Study

Observation ratings should be treated as **one source of evidence** regarding learner experiences. The rubric descriptors (e.g., *Pre-Emergent*, *Baseline*, *Maturing*, and *Advanced*) should be interpreted as **analytic descriptors or categories for organizing and synthesizing evidence**, not as validated performance levels or classifications of classrooms, schools, or educators.

For the remaining analyses and case studies, classroom observations should be integrated with student and teacher interviews, surveys, artifacts, and document review to develop evidence-based interpretations of each CBL Marker. Findings should emphasize the **convergence and divergence of evidence across methods**, instead of relying on classroom observation ratings alone. Similarly, observation findings should be used to provide contextual evidence that informs interpretation of broader patterns observed across qualitative and quantitative data sources.

Use Guidance

The observation protocol is intended to support systematic observation and documentation of learner behaviors associated with competency-based learning. It should be used as one component of a broader evidence collection strategy that includes complementary sources of evidence (e.g., interviews, artifacts, and document review). It should not be used as a stand-alone measure of CBL. Ongoing field testing should continue to refine the behavioral indicators, scoring guidance, annotations, and evidence collection procedures.

Appendix D.

Psychometric Analysis of the CBL Student Survey

Purpose and Research Question

The CBL Student Survey was developed to examine learners' perceptions of competency-based learning (CBL) across participating high schools. This appendix examines the measurement quality evidence supporting interpretation of survey responses, including item functioning, internal consistency, the hypothesized factor structure, empirical distinction among closely related CBL markers, and relationships with theoretically related motivational and instructional constructs.

Consistent with contemporary approaches to validation, validity is treated as the degree to which theory and empirical evidence support the intended interpretations of scores for their proposed uses (American Educational Research Association [AERA] et al., 2014; Kane, 2013).

The following research question guided the analysis:

- RQ2. What evidence supports interpretations and inferences drawn from the student survey about learners' experiences of competency-based learning?

The analyses addressed two complementary aspects of this question: whether the proposed marker scales demonstrated acceptable item functioning and a distinguishable internal structure, and whether learners' reported CBL experiences showed theoretically expected relationships with mastery and performance-approach goal orientations, personalization, and project-based learning.

Appendix B describes the development and definitions of the ten hypothesized CBL markers. The complete Student Survey appears in *Appendix I*.

Methods

Study Sample and Administration

Learners represented all four high school grade levels, with somewhat greater representation in grades 9 and 10 (Table D1). Nearly half identified as male (47.6%) and 44.0% as female, and approximately one in five (19.0%) reported receiving IEP or Section 504 supports. The sample was racially and ethnically diverse: 35.9% identified as White, 28.4% as Hispanic or Latino, 13.8% as Black or African American, and 12.2% selected more than one race/ethnicity category. Nearly half (46.8%) reported earning mostly A's during the prior year, although 7.9% attended high schools that did not use grades.

The initial psychometric analyses used the 1,328 survey responses available at the time when the item-level and internal-consistency analyses were conducted. Additional learner responses were received at a later date, increasing the available sample to 1,350 learners from ten participating high schools. The confirmatory factor analyses and analyses involving external measures used this expanded sample.

Table D1. Learner Reported Demographics

Learner Reported Characteristic	Category	<i>n</i>	%
Grade (<i>n</i> = 1,297)	9th grade	419	32.0
	10th grade	341	26.0
	11th grade	318	24.3
	12th grade	219	16.7
Gender (<i>n</i> = 1,302)	Female	573	44.0
	Male	620	47.6
	Other gender	46	3.5
	Prefer not to say	63	4.8
Years at high school (<i>n</i> = 1,307)	Less than 1 year	327	25.0
	1–2 years	384	29.4
	3–4 years	462	35.3
	More than 4 years	134	10.3
IEP or 504 support (<i>n</i> = 1,302)	Yes	248	19.0
	No	770	59.1
	Not sure	212	16.3
	Prefer not to say	72	5.5
Grades in prior year (<i>n</i> = 1,313)	Mostly A's	615	46.8
	Mostly B's	321	24.4
	Mostly C's	61	4.6

	Mostly D's or F's	16	1.2
	Not sure	114	8.7
	Prefer not to say	82	6.2
	High school does not use grades	104	7.9
Race/ethnicity ($n = 1,308$; select all that apply)	White or Caucasian	469	35.9
	Hispanic or Latino	372	28.4
	Black or African American	180	13.8
	Asian	28	2.1
	American Indian or Alaskan Native	17	1.3
	Multiracial or Biracial	14	1.1
	A race not listed here	6	0.5
	Middle Eastern or North African	4	0.3
	Native American or Alaskan Native	1	0.1
	Prefer not to say	57	4.4
	Selected more than one race/ethnicity category	160	12.2

Note. Learners could self-select to report demographic information, resulting in a smaller sample reflected in the table.

Accordingly, item-functioning and internal-consistency results reported in this appendix reflect the initial analytic dataset, while the CFA and latent-variable analyses reflect the updated sample. Sample sizes within individual analyses also vary because of item-level missingness and the analytic requirements of particular measures.

The survey was administered online through Qualtrics during the spring 2026 study data-collection period and required approximately 15–20 minutes to complete. Appendix A describes the broader sampling, recruitment, administration, and participant-protection procedures.

Instrument

The proposed CBL markers were conceptualized as latent constructs representing related but distinguishable dimensions of learners' experiences. The Student Survey operationalized those constructs through multiple items measuring learners' perceptions of how competencies organize expectations, learning, assessment, progression, and evidence of mastery.

The instrument was developed to capture competency-based learning from the learner perspective and to parallel the conceptual framework represented in the study's other measures. Appendix B describes the literature synthesis, expert review, and iterative development process used to define the marker set.

The initial CBL Student Survey contained 35 CBL items representing ten proposed markers: *Clear Articulation* (5 items), *Developmental Progression* (4 items), *Connection to Learning Targets* (3 items), *Universal Application of Expectations* (4 items), *Demonstration Location* (3 items), *Clear Calibration* (3 items), *Rigorous Application of Learning* (3 items), *Mastery-Based Progression* (4 items), *Iterative Purpose* (3 items), and *Transparency* (3 items).

Items used five-point Likert-type response scales. Depending on item content, response options assessed frequency (*Never to Always*), agreement (*Strongly disagree to Strongly agree*), or perceived truthfulness (*Not at all true to Very true*). Higher scores represented stronger learner reports of the corresponding CBL experience.

Preliminary item-level and internal-consistency analyses identified insufficient coherence among the three *Demonstration Location* items, labeled *Any Location* in the original analytic files. Those items were not included in the primary CFA. The resulting measurement model included nine CBL factors represented by 32 indicators.

External Measures

Four measures were included to examine relationships between learners' reported CBL experiences and theoretically related constructs.

- **Mastery Goal Orientation** was assessed using five items adapted from the *Patterns of Adaptive Learning Scales* (PALS; Midgley et al., 2000). The items measured learners' emphasis on learning, understanding, skill development, and improvement. Responses used a five-point scale ranging from *Not at all true* to *Very true*, with higher scores indicating stronger mastery goal orientation. Mastery Goal Orientation scores were expected to be positively correlated with CBL Marker responses, such that students with stronger mastery-goal orientation would report more frequent or stronger CBL experiences.

- **Performance Goal Orientation** was assessed using five items adapted from PALS (Midgley et al., 2000). These items focused on demonstrating competence to others and appearing capable relative to classmates. Responses used the same five-point truthfulness scale. This construct was labeled *Performance Goal Orientation* in some analytic outputs. *Performance Goal Orientation* scores were expected to be weakly correlated with CBL Marker responses, such that students with higher performance-approach goal orientation would report less frequent or stronger CBL experiences.
- **Personalization** was assessed using four items adapted from the Personalized Instruction domains of the Education Policy Innovation Collaborative's competency-based education Student Survey (EPIC, 2021). Two items came from the Personalized Content domain and two from the Personalized Rigor domain. Items were adapted from class-specific to broader high school-learning contexts and used a five-point frequency scale ranging from *Never* to *Always*. Personalization scores were expected to be moderately and positively correlated with CBL Marker responses, such that students reporting more frequent personalization would also report moderate to strong CBL experiences.
- **Project-Based Learning** was assessed using seven items adapted from the EPIC (2021) Student Survey. Items addressed project-centered instruction, application of learning across contexts, and the role of projects in learning and assessment. Responses used a five-point frequency scale ranging from *Never* to *Always*. Project-based Learning scores were expected to be moderately and positively correlated with CBL Marker responses, such that students reporting more frequent project-based learning experiences would also report moderate to strong CBL experiences.

Item Analysis and Internal Consistency

Item-level analyses were conducted in R version 2025.05.1+513 to examine missingness, response frequencies, variability, item means, and corrected item-total correlations. These analyses assessed whether items provided sufficient variation and functioned coherently within their proposed constructs.

Internal consistency was evaluated using Cronbach's coefficient alpha for each proposed CBL marker and each external measure (Cronbach, 1951). Alpha-if-item-deleted results were also examined to identify items whose removal might improve scale coherence. Reliability estimates were interpreted alongside item functioning and subsequent CFA results.

Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) evaluated the hypothesized correlated nine-factor structure of the Student Survey. Each of the 32 observed indicators was specified to load only on its theoretically assigned latent factor. Cross-loadings and residual covariances were fixed to zero, all nine latent factors were allowed to correlate, and latent factor variances were fixed to one for model identification.

Given the five-category response formats, low item-level missingness, lack of sparsely populated response categories, and modest distributional asymmetry, indicators were treated as approximately continuous. Models were estimated using robust maximum likelihood (MLR), which provides standard errors and test statistics robust to departures from multivariate normality. Simulation research indicates that five-category indicators can often be treated as approximately continuous when distributions are not severely nonnormal (Rhemtulla et al., 2012).

Missing responses were handled using full-information maximum likelihood (FIML), allowing learners with partially observed indicators to contribute available information to model estimation. FIML can provide appropriate estimates under suitable missing-data assumptions; low missingness itself does not establish that those assumptions hold (Enders & Bandalos, 2001).

Models were estimated in R using the lavaan package (Rosseel, 2012). Overall fit was evaluated using the model chi-square statistic, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA) with its 90% confidence interval, and standardized root mean square residual (SRMR). Values near .95 for CFI and TLI, near .06 for RMSEA, and .08 or lower for SRMR served as descriptive reference points rather than rigid pass/fail criteria (Hu & Bentler, 1999).

Model fit was interpreted alongside standardized factor loadings, latent correlations, residual diagnostics, and solution admissibility. Loadings below .40 were treated as warranting substantive review. Standardized residuals and modification indices were examined to identify localized sources of model misfit. No post hoc cross-loadings, residual covariances, or other parameters were added to the reported model.

Discriminant Validity Among CBL Markers

Latent factor correlations were examined to assess overlap among the nine CBL markers. Factor pairs with correlations of .80 or higher were flagged for closer evaluation rather than treated automatically as evidence of construct redundancy (Brown, 2015; Rönkkö & Cho, 2022).

For each flagged pair, the latent correlation was constrained to 1.00 in an otherwise equivalent model. The constrained model was compared with the freely estimated model using a Satorra-Bentler scaled chi-square difference test. A statistically significant deterioration in model fit was interpreted as evidence that the factors were statistically distinguishable under this test despite substantial shared variance (Jöreskog, 1971; Rönkkö & Cho, 2022).

Relationships With External Measures

To examine whether CBL scores showed theoretically expected relationships with neighboring constructs, the nine CBL factors were estimated simultaneously with Mastery Goal Orientation, Performance-Approach Goal Orientation, Personalization, and Project-Based Learning in a 13-factor MLR model. The same general estimation and model-evaluation procedures used for the primary CFA were applied.

The external constructs were selected a priori based on the conceptual framework. CBL dimensions were expected to be positively associated with mastery goal orientation, personalization, and project-based learning and to show comparatively weaker associations with performance-approach goal orientation. Latent correlations were examined to assess whether the observed pattern was consistent with these expectations. Formal discriminant-validity tests between the CBL factors and external constructs were not conducted.

Findings

Item Functioning

Across the CBL items, means ranged from 2.81 to 3.72 on the five-point response scales, and standard deviations ranged from 0.86 to 1.25. Item nonresponse was consistently low, ranging from approximately 1% to 3%. Table D2 summarizes item functioning for the ten proposed markers and four external measures.

Table D2. Summary of Student Survey Item Analysis Results

Construct	Items	Missing responses (%)	Item mean range	Corrected item-total correlation range
CBL markers				
Clear Articulation	5	1.4–2.6	2.84–3.60	.48–.77
Developmental Progression	4	1.7–2.1	3.30–3.39	.69–.74
Connection to Learning Targets	3	2.3–2.9	3.37–3.60	.69–.76
Universal Application of Expectations	4	1.9–2.9	3.25–3.68	.55–.67

Demonstration Location	3	1.8–2.5	3.11–3.29	–.25–.20
Clear Calibration	3	1.9–2.1	3.43–3.49	.40–.60
Rigorous Application of Learning	3	1.9–2.4	3.36–3.50	.65–.68
Mastery-Based Progression	4	2.0–2.5	3.53–3.71	.64–.76
Iterative Purpose	3	2.2–2.6	3.48–3.72	.61–.68
Transparency	3	2.3–2.4	2.81–2.96	.76–.80
External measures				
Mastery Goal Orientation	5	1.8–2.3	3.73–3.98	.71–.78
Performance-Approach Goal Orientation	5	2.1–2.6	2.75–3.30	.65–.82
Personalization	4	2.2–2.6	3.00–3.44	.48–.67
Project-Based Learning	7	2.1–2.8	2.45–3.87	.26–.66

Note. Ranges describe variation across items within each construct. *Demonstration Location* was labeled Any Location in the original analyses. Marker names have been standardized to the terminology used in this report.

Nine of the ten proposed CBL marker sets demonstrated generally coherent item functioning. Corrected item-total correlations ranged from approximately .40 to .80 across those markers. Alpha-if-item-deleted results did not suggest that removing individual items from those scales would materially improve internal consistency.

Demonstration Location was the principal exception. The reverse-coded item assessing whether only school personnel could determine mastery had a negative corrected item-total correlation of –.25, while the remaining two items had weak correlations of .19 and .20. The three items therefore did not function coherently as indicators of a single construct.

One possible interpretation is that learners distinguish among different aspects of demonstrating learning across settings, including who can determine mastery, whether evidence generated beyond the high school counts toward mastery, and whether people outside the high school can evaluate competency. The current analysis cannot establish that explanation, but it identifies a need for additional conceptual and item-level refinement.

The external measures also demonstrated generally coherent item functioning. One Project-Based Learning item—whether projects completed in one class also counted toward grades in other classes—had a comparatively weak corrected item-total correlation of .26.

Internal Consistency

Cronbach's coefficient alpha ranged from .71 to .88 for nine of the ten proposed CBL markers. Transparency showed the highest internal consistency, followed by Clear Articulation, Developmental Progression, and Mastery-Based Progression. The four external measures had alpha coefficients ranging from .79 to .90 (Table D3).

Table D3. Internal Consistency Reliability of Student Survey Scales

Construct	Analytic n	Items	Cronbach's α
CBL markers			
Clear Articulation	1,285	5	.87
Developmental Progression	1,296	4	.86
Connection to Learning Targets	1,284	3	.85
Universal Application of Expectations	1,289	4	.79
Demonstration Location	1,294	3	.02
Clear Calibration	1,289	3	.71
Rigorous Application of Learning	1,294	3	.81
Mastery-Based Progression	1,286	4	.86
Iterative Purpose	1,290	3	.79
Transparency	1,291	3	.88
External measures			
Mastery Goal Orientation	1,287	5	.90
Performance-Approach Goal Orientation	1,288	5	.90
Personalization	1,286	4	.79
Project-Based Learning	1,275	7	.80

Note. These estimates come from the initial psychometric-analysis dataset of 1,328 responses. Analytic n represents the number of respondents contributing to each scale-specific reliability estimate and varies because of item nonresponse. Subsequent CFA analyses used the expanded sample of 1,350 responses.

Consistent with the item analysis, *Demonstration Location* had extremely low internal consistency ($\alpha = .02$) and was not supported as a composite scale. The three *Demonstration Location* items were therefore excluded from the primary CFA. Individual items may still be examined descriptively while this portion of the Student Survey is refined. This finding concerns the functioning of the current items; it does not determine whether the underlying learner experience is conceptually meaningful within the broader marker framework.

The external measures demonstrated satisfactory internal consistency and were retained for subsequent analyses examining their relationships with the CBL factors.

Overall Fit of the Nine-Factor Model

The primary CFA used the expanded sample of 1,350 learners and included the 32 indicators associated with the nine CBL markers retained after preliminary analysis.

The hypothesized nine-factor model converged successfully and yielded an admissible solution. There were no negative observed-item residual variances, no standardized factor loadings greater than 1, and no latent factor correlations exceeding 1.

As shown in Table D4, the model chi-square was statistically significant, $\chi^2(428) = 1,823.70$, $p < .001$. RMSEA and SRMR indicated favorable approximate fit, while CFI and TLI fell below the .95 reference point.

Table D4. Fit Indices for the Nine-Factor Student CBL Model

Fit statistic	Estimate
Analytic N	1,350
χ^2	1,823.70
Degrees of freedom	428
p	< .001
CFI	.923
TLI	.910
RMSEA	.049

RMSEA 90% CI	[.047, .051]
SRMR	.049

Note. The model included nine correlated factors and 32 indicators and was estimated using robust maximum likelihood. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Taken together, these findings provide qualified support for the hypothesized nine-factor structure. RMSEA and SRMR indicated favorable approximate fit, whereas CFI and TLI remained below the stated reference values. The model therefore captured substantial structure in the data without fully reproducing all observed relationships among the items.

Standardized Factor Loadings

All 32 indicators loaded positively and significantly on their hypothesized factors ($p < .001$). Standardized loadings, shown in Table D5, ranged from .562 to .881, and none fell below the .40 review threshold.

Table D5. Standardized Factor Loadings for the Nine-Factor Student CBL Model

Marker	Indicator	Standardized Loading	95% CI
Clear Articulation	Use competencies to plan work across classes	.813	[.786, .840]
	Use competencies to plan learning outside classes	.805	[.775, .835]
	Recognize competencies across classes	.833	[.809, .858]
	Recognize competencies outside school	.800	[.773, .827]
	Competencies prepare me for my future	.562	[.511, .613]
Developmental Progression	Choose experiences that support higher competency levels	.775	[.740, .810]
	Explain improvement at a higher competency level	.768	[.736, .801]
	Explain the competency level I am working toward	.804	[.773, .835]
	Explain how close I am to mastery	.791	[.754, .827]

Connection to Learning Targets	Explain concepts and skills being worked on	.784	[.751, .817]
	Explain closeness to mastery in classes	.858	[.829, .886]
	Explain requirements for each level	.786	[.754, .818]
Universal Application of Expectations	Use rubric to connect skills to a broader competency	.774	[.735, .812]
	Use rubric describing mastery criteria	.797	[.759, .834]
	Expect adults to use the same evaluation criteria	.618	[.561, .675]
	Same criteria used across learning experiences	.614	[.555, .672]
Clear Calibration	Results consistent regardless of who assesses	.591	[.502, .679]
	Results consistent regardless of how assessed	.619	[.535, .703]
	Use assessment results to determine next learning steps	.736	[.687, .784]
Rigorous Application of Learning	Apply learning to authentic problems or situations	.764	[.728, .800]
	Demonstrate mastery through meaningful performance tasks	.797	[.764, .830]
	Combine skills to address authentic problems	.756	[.717, .796]
Mastery-Based Progression	Continue working until mastery	.805	[.773, .838]
	Move on when ready after demonstrating mastery	.703	[.664, .742]
	Continue working even when additional time is needed	.826	[.797, .854]
	Improve work until mastery criteria are met	.809	[.780, .838]
Iterative Purpose	Revise/practice before receiving final credit	.739	[.703, .775]

	Retake assessments when needed	.704	[.662, .747]
	Use multiple drafts or assessment attempts	.811	[.782, .840]
Transparency	Maintain tracker/portfolio organized by competencies	.835	[.807, .863]
	Use tracker/portfolio to monitor progress	.881	[.856, .906]
	Share tracker/portfolio to obtain feedback	.824	[.797, .852]

Note. Indicator descriptions are abbreviated analytic labels rather than the complete administered item wording. All loadings were statistically significant at $p < .001$. The complete Student Survey appears in Appendix I.

Transparency showed particularly strong and consistent indicator loadings. *Developmental Progression*, *Connection to Learning Targets*, *Mastery-Based Progression*, *Rigorous Application of Learning*, and *Iterative Purpose* also showed consistently strong loadings.

Indicator strength was somewhat more variable for *Universal Application of Expectations* and *Clear Calibration*. The weakest loading was the *Clear Articulation* item concerning whether competencies would prepare learners for their future, $\lambda = .562$, 95% CI [.511, .613]. Although weaker than the other *Clear Articulation* indicators, the loading remained meaningful and statistically significant. Table D6 summarizes indicator strength within each factor.

Table D6. Summary of Indicator Strength by CBL Marker

Marker	Items	Mean loading	Loading range
Clear Articulation	5	.763	.562–.833
Developmental Progression	4	.784	.768–.804
Connection to Learning Targets	3	.809	.784–.858
Universal Application of Expectations	4	.701	.614–.797
Clear Calibration	3	.649	.591–.736
Rigorous Application of Learning	3	.772	.756–.797
Mastery-Based Progression	4	.786	.703–.826
Iterative Purpose	3	.751	.704–.811
Transparency	3	.847	.824–.881

Note. Mean loadings summarize the standardized indicator loadings within each factor; they are not separate reliability coefficients.

Relationships Among the CBL Factors

All nine CBL factors were positively correlated, as shown in Table D7, with latent correlations ranging from .488 to .874. The strongest association was between *Mastery-Based Progression* and *Iterative Purpose*, followed by *Developmental Progression* and *Connection to Learning Targets*.

Table D7. Latent Correlations Among the Nine Student CBL Factors

Marker	CA	DP	LT	UE	CC	RA	MP	IP	TR
CA	—								
DP	.845	—							
LT	.640	.852	—						
UE	.707	.816	.755	—					
CC	.658	.749	.722	.761	—				
RA	.674	.735	.690	.745	.844	—			
MP	.615	.714	.672	.705	.822	.684	—		
IP	.656	.732	.693	.714	.819	.722	.874	—	
TR	.652	.599	.488	.564	.516	.535	.516	.586	—

Note. CA = Clear Articulation; DP = Developmental Progression; LT = Connection to Learning Targets; UE = Universal Application of Expectations; CC = Clear Calibration; RA = Rigorous Application of Learning; MP = Mastery-Based Progression; IP = Iterative Purpose; TR = Transparency. Values are model-estimated latent correlations.

The strength of several relationships indicates substantial shared variance among conceptually adjacent CBL experiences. These correlations do not, by themselves, establish whether the corresponding markers should be combined.

Discriminant Validity Among CBL Markers

Seven factor pairs had latent correlations of .80 or higher and were evaluated using Satorra-Bentler scaled chi-square difference tests. As shown in Table D8, for all seven pairs, constraining the latent correlation to 1.00 significantly worsened model fit ($p < .001$).

Table D8. Discriminant Validity Tests for Highly Correlated CBL Factor Pairs

Factor pair	Latent r	Scale d $\Delta\chi^2$	Δdf	p	Interpretation
Mastery-Based Progression–Iterative Purpose	.874	52.03	1	< .001	Statistically distinguishable
Developmental Progression–Connection to Learning Targets	.852	89.24	1	< .001	Statistically distinguishable
Clear Articulation–Developmental Progression	.845	163.15	1	< .001	Statistically distinguishable
Clear Calibration–Rigorous Application of Learning	.844	19.75	1	< .001	Statistically distinguishable
Clear Calibration–Mastery-Based Progression	.822	27.46	1	< .001	Statistically distinguishable
Clear Calibration–Iterative Purpose	.819	23.56	1	< .001	Statistically distinguishable
Developmental Progression–Universal Application of Expectations	.816	96.03	1	< .001	Statistically distinguishable

Note. Each test compared the freely estimated primary model with an otherwise equivalent model constraining the specified factor correlation to 1.00. “Statistically distinguishable” indicates that the unity constraint produced significantly worse fit. It does not imply weak association or fully resolved conceptual boundaries.

These findings support retaining the nine markers separately in the reported student measurement model despite substantial overlap among several dimensions. Clear Calibration appeared in three of the seven highly correlated pairs. Transparency did not reach the .80 testing threshold; its relationships with the other CBL factors were comparatively moderate.

Local Model Diagnostics

Standardized residuals and modification indices identified several localized item relationships not fully captured by the hypothesized simple-structure model.

The largest standardized residual involved the two Clear Calibration indicators assessing whether mastery results were consistent regardless of who assessed the work and how the work was assessed (residual = .274). Their parallel wording and closely related content suggest shared item-specific variance beyond that represented by the Clear Calibration factor.

The Clear Articulation item concerning whether competencies would “help prepare me for my future” also appeared repeatedly among larger residual relationships and modification indices. Together with its comparatively weaker factor loading, this pattern suggests that learners may interpret future preparation as a broader feature of their learning experience rather than solely as an indicator of Clear Articulation.

A further diagnostic involved the Clear Calibration item assessing use of assessment results to determine next learning steps. Although the item loaded meaningfully on Clear Calibration ($\lambda = .736$), modification indices suggested additional relationships with Iterative Purpose and Mastery-Based Progression. Conceptually, the item may bridge assessment calibration and progression because it explicitly connects assessment evidence to decisions about revising, improving, or moving forward.

These diagnostics identify priorities for future item refinement. No residual covariances, cross-loadings, or other post hoc parameters were added to the reported model.

Relationships With External Measures

The 13-factor model estimated the nine CBL factors together with *Mastery Goal Orientation*, *Performance-Approach Goal Orientation*, and *Project-Based Learning*. The model converged successfully and produced an admissible solution.

RMSEA and SRMR indicated favorable approximate fit, while CFI and TLI fell below the .95 reference point (Table D9). The pattern and magnitude of the CBL factor loadings remained similar to those in the primary nine-factor model.

Table D9. Fit Indices for the Student CBL Model With External Measures

Fit statistic	Estimate
Analytic <i>N</i>	1,350
χ^2	4,215.21
Degrees of freedom	1,247
<i>p</i>	< .001
CFI	.906

TLI	.896
RMSEA	.042
RMSEA 90% CI	[.041, .043]
SRMR	.060

Note. The model included nine CBL factors and four external factors represented by 53 indicators and was estimated using robust maximum likelihood. Because it included additional measures, this model was not a nested comparison with the nine-factor CFA.

The external constructs were generally well represented by their indicators. Standardized loadings ranged from .763 to .834 for Mastery Goal Orientation, .663 to .879 for Performance-Approach Goal Orientation, and .592 to .779 for Personalization. Project-Based Learning loadings were more variable, ranging from .313 to .758. Consistent with the earlier item analysis, the item assessing whether projects completed in one class also counted toward grades in other classes had the weakest loading.

CBL markers were moderately to strongly associated with Mastery Goal Orientation, with correlations ranging from .382 to .698, and showed consistently weaker associations with Performance-Approach Goal Orientation, ranging from .310 to .417. For each of the nine CBL markers, the estimated correlation with mastery goals exceeded the corresponding correlation with performance-approach goals.

This pattern is consistent with the theoretical expectation that learners' reported CBL experiences would relate more closely to goals emphasizing learning and improvement than to goals emphasizing demonstration of ability relative to others. Formal statistical comparisons of the paired correlation magnitudes were not conducted, however, so the pattern should be interpreted descriptively.

CBL markers were also positively associated with Personalization (.537-.720) and Project-Based Learning (.482-.800). Relationships with Project-Based Learning were particularly strong for Clear Calibration ($r = .800$) and Rigorous Application of Learning ($r = .786$).

These associations provide evidence that the CBL Student Survey relates to theoretically relevant neighboring constructs in expected ways while also showing substantial overlap with related learner-centered experiences.

Interpretation and Analytic Limitations

The CBL Student Survey analyses provide generally supportive but appropriately qualified evidence for marker-level interpretation.

Nine proposed CBL marker scales demonstrated coherent item functioning and internal consistency in the initial psychometric analyses. In the updated CFA sample, all 32 retained indicators loaded positively and significantly on their hypothesized factors, and the nine-factor model produced an admissible solution. Seven highly correlated factor pairs were also statistically distinguishable from perfect correlation under the targeted Satorra-Bentler tests.

At the same time, the overall fit evidence was mixed. RMSEA and SRMR indicated favorable approximate fit, while CFI and TLI remained below the prespecified .95 reference values. Several latent correlations were high, and local diagnostics identified items whose content may span conceptual boundaries. The evidence therefore supports retaining the nine-factor structure for this study while recognizing meaningful overlap among several dimensions and opportunities for further item refinement.

The *Demonstration Location* items present a separate measurement limitation. Their weak item-total correlations and very low internal consistency did not support construction of a marker-level scale or inclusion as a latent factor in the CFA. This finding concerns the current survey operationalization and does not determine whether Demonstration Location remains conceptually meaningful within the broader marker framework.

Measurement strength should also be distinguished from the level of experience learners reported. Transparency, for example, demonstrated strong item functioning, internal consistency, and factor loadings even though its substantive scores were comparatively low. Lower average ratings therefore should not be interpreted as evidence that the measure itself performed poorly.

Relationships with external constructs were generally consistent with theoretical expectations. CBL factors were more strongly associated descriptively with mastery than with performance-approach goal orientation and were positively associated with Personalization and Project-Based Learning. These relationships support the theoretical positioning of the Student Survey but do not, in the absence of formal external discriminant-validity testing, establish that all CBL dimensions are empirically distinct from these neighboring constructs.

The analyses are based on learner self-reports from ten participating high schools. They do not establish that the measurement structure operates equivalently across high schools, demographic groups, or time; measurement-invariance and longitudinal analyses were not conducted. Relationships among constructs measured within the same survey may also reflect shared response processes. The analyses do not test causal relationships between learners' CBL experiences and motivational orientations, instructional approaches, or academic outcomes.

Taken together, the findings support use of the nine retained CBL Student Survey scales as descriptive measures of learners' reported experiences associated with individual CBL markers, with the qualifications described above. They do not support a single overall CBL score, treatment of Demonstration Location as a reliable three-item composite, or causal interpretation of relationships among measured experiences.

Future research can refine the identified items, test the proposed structure in additional learner samples, examine measurement equivalence across populations and settings, and investigate whether score interpretations remain stable over time.

Appendix E.

Psychometric Analysis of the CBL Teacher Survey

Purpose and Research Question

The CBL Teacher Survey provided complementary evidence about educators' perceptions of learners' experiences with competencies and the instructional practices associated with those experiences. This appendix examines the measurement evidence supporting interpretation of teacher responses, including item functioning, internal consistency, the hypothesized factor structure, and the empirical distinction among closely related markers.

The following research question guided the analysis:

- RQ3. What evidence supports interpretations and inferences drawn from the teacher survey about learners' experiences of competency-based learning?

Appendix B describes the development and definitions of the ten hypothesized CBL markers. The complete teacher survey instrument is included in *Appendix J*.

Methods

Study Sample and Administration

The initial psychometric analyses used the 139 teacher responses available when those analyses were conducted. Two additional teacher responses were received subsequently, increasing the available sample to 141 respondents from ten participating high schools.

As shown in E1, the teacher sample represented a range of experience levels and instructional roles. Nearly half of respondents (48.9%) reported 15 or more years of teaching experience overall, while experience at their current high school was more varied, with 40.0% reporting 4–9 years and 26.4% reporting 3 years or less. Respondents also represented a broad mix of teaching areas, most commonly English language arts, mathematics, science, and social studies. Because teaching area was a select-all-that-apply item, percentages across subject areas exceed 100%.

Table E1. Teacher Reported Demographics

Characteristic	Category	<i>n</i>	%
Years teaching at this high school (<i>n</i> = 140)	3 years or less	37	26.4
	4–9 years	56	40.0
	10–14 years	21	15.0
	15 or more years	26	18.6
Total years teaching (<i>n</i> = 141)	3 years or less	7	5.0
	4–9 years	32	22.7
	10–14 years	33	23.4
	15 or more years	69	48.9
Primary area(s) of teaching/facilitating learning (<i>n</i> = 141; select all that apply)	English Language Arts	34	24.1
	Mathematics	28	19.9
	Science	23	16.3
	Social Studies	22	15.6
	The Arts	16	11.3
	Career/Technical Education	13	9.2
	Foreign Language	13	9.2
	Special Education	12	8.5
	Other	11	7.8
	Physical Education	6	4.3
	Instructional Coach	2	1.4

Note. Teachers could self-select to report demographic information, resulting in a smaller sample reflected in the table.

Item-level analyses and internal consistency estimates reported in this appendix reflect the initial sample. Confirmatory factor analysis and subsequent discriminant-validity analyses used the expanded sample.

The survey was administered during the spring 2026 study data-collection period. *Appendix A* describes the broader sampling, recruitment, administration, and participant-protection procedures.

Instrument

The proposed competency-based learning markers were conceptualized as latent constructs representing coherent patterns of CBL practices. The Teacher Survey operationalized each construct or marker through multiple items measuring educators' perceptions of students' engagement with competency structures, such as competencies, mastery criteria, and progressions, and the instructional practices, such as feedback, revision, and evidence-based progression, that support those interactions. The instrument was developed to parallel the conceptual framework represented in the CBL Observation Protocol, providing complementary evidence from educators about the extent to which learners experience the hypothesized markers of competency-based learning.

The conceptual underpinnings of the survey were grounded in the hypothesis that competency-based learning makes expectations, progression, assessment, and evidence of learning increasingly coherent, transparent, and consequential in learners' academic and lived experiences. Consistent with the broader measurement framework, the survey was designed to capture the extent to which learners understand and use competency structures, act on evidence of their learning, and experience opportunities to progress, revise, and demonstrate learning in relation to defined expectations. Higher scores therefore represented stronger educator perceptions of the corresponding CBL experience or practice.

Items were developed to operationalize the ten proposed competency-based learning markers: *Clear Articulation*, *Developmental Progression*, *Connection to Learning Targets*, *Universal Application of Expectations*, *Demonstration Location*, *Clear Calibration*, *Rigorous Application of Learning*, *Mastery-Based Progression*, *Iterative Purpose*, and *Transparency*. Each item was written to reflect learner experiences or instructional practices hypothesized to make competency-based learning visible.

The survey comprised four sections organized around major dimensions of competency-based learning: *Learning Expectations in Your School*, *Assessment in Your School*, *Understanding Progression in Your School*, and *Understanding Growth in Your School*. The Learning Expectations section included items measuring *Clear Articulation* (5 items), *Developmental Progression* (4 items), *Connection to Learning Targets* (4 items), and *Universal Application of Expectations* (3 items). The Assessment section included items measuring *Demonstration Location* (3 items), *Clear Calibration* (3 items), and *Rigorous Application of Learning* (3 items). The Understanding Progression section included items measuring *Mastery-Based Progression* (4 items) and *Iterative Purpose* (3 items).

The Understanding Growth section included items measuring *Transparency* (3 items). In total, the instrument included 35 items representing the ten proposed CBL markers.

Respondents indicated the extent to which each statement described their classroom or high school using five-point Likert-type response scales, such as *Never* to *Always*, *Strongly Disagree* to *Strongly Agree*, or *Not at all true* to *Very true*, depending on the nature of the item.

The primary confirmatory factor analysis included nine markers represented by 32 observed indicators. The three-item *Demonstration Location* scale was excluded because preliminary item-level and internal consistency analyses indicated weak coherence among its indicators.

Item Analysis and Internal Consistency

Item-level analyses were conducted in R to examine item functioning and identify potential threats to measurement quality, including limited response variability, weak item discrimination, and missingness. Corrected item-total correlations were examined to assess how strongly each item related to the remaining items within its proposed scale.

Cronbach's alpha was calculated to estimate internal consistency within each marker scale (Cronbach, 1951). These analyses provided preliminary evidence about whether items within each proposed marker functioned coherently. Internal consistency alone does not establish that a scale measures the intended construct or that proposed markers are empirically distinct.

Confirmatory Factor Analysis

A confirmatory factor analysis (CFA) was conducted to evaluate the hypothesized correlated nine-factor structure of the teacher measure. Each of the 32 observed indicators was specified to load only on its theoretically assigned latent factor. Cross-loadings and residual covariances were fixed to zero, all nine latent factors were allowed to correlate, and latent factor variances were fixed to 1 for model identification.

Given the five-category response formats, relatively small teacher sample, and sparsely populated response categories for some items, indicators were treated as approximately continuous, and the model was estimated using robust maximum likelihood (MLR). MLR provides standard errors and test statistics that are robust to departures from multivariate normality. Continuous-variable estimation methods can perform adequately with five-category indicators under suitable distributional conditions, although performance depends on the characteristics of the data (Rhemtulla et al., 2012).

Missing item responses were handled using full-information maximum likelihood (FIML), allowing respondents with partially observed indicators to contribute available information to model estimation.

The CFA was estimated in R using the lavaan package (Rosseel, 2012). Model fit was evaluated using the robust chi-square statistic, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), root mean square error of approximation (RMSEA) with its 90% confidence interval, and standardized root mean square residual (SRMR). CFI and TLI values near .95, RMSEA near .06, and SRMR at or below .08 served as interpretive reference points rather than absolute decision rules (Hu & Bentler, 1999).

Global model fit was interpreted alongside standardized factor loadings, latent correlations, residual diagnostics, and solution admissibility. Standardized factor loadings below .40 were identified as indicators warranting substantive review. Standardized residuals and modification indices were inspected diagnostically for localized areas of model misfit. No cross-loadings, residual covariances, or other post hoc parameters were added to the reported model.

Discriminant Validity Analysis

Latent factor correlations exceeding .80 in the primary model prompted targeted follow-up analyses. For each selected factor pair, a model allowing the correlation to be freely estimated was compared with an otherwise equivalent model constraining that correlation to 1.00. Models were compared using a Satorra–Bentler scaled chi-square difference test.

A statistically significant difference indicates that constraining the factors to correlate perfectly worsens model fit, providing evidence against treating them as empirically identical in that comparison. These tests contribute evidence about empirical distinction but do not establish discriminant validity comprehensively (Rönkkö & Cho, 2022).

Findings

Missingness

Missingness varies by subconstruct, as shown in the table below. There is a fairly monotonic increase in missingness as respondents progress through the survey. Item nonresponse increased gradually with survey position. Although two constructs (*Clear Articulation* and *Clear Calibration*) included items administered later than the remainder of their scales, those items exhibited missingness levels consistent with their position in the survey rather than with the overall pattern for their constructs. This pattern suggests that survey attrition, rather than the content of specific competency-based learning constructs, was the primary driver of item nonresponse.

In addition to investigating ways to reduce survey length without jeopardizing construct validity, future survey administrations should randomize the presentations of each section to ensure no one section is systematically impacted by survey fatigue and to increase retention.

Table E2. Missingness Diagnostic Analysis for CBL Markers Teacher Survey

Survey Section	CBL Marker	Total Responses	Complete Cases	% Missing	Approximate Survey Position
Learning Expectations	Developmental Progression	139	137	1%	1
Learning Expectations	Universal Expectations	139	135	3%	2
Learning Expectations	Learning Targets	139	135	3%	3
Learning Expectations (end)	Clear Articulation*	139	134	4%	4
Assessment	Demonstration Location	139	133	4%	5
Assessment	Rigorous Application	139	131	6%	6
Understanding Progression	Mastery-Based Progression	139	131	6%	7
Understanding Progression	Iterative Purpose	139	130	6%	8
Understanding Progression (end)	Clear Calibration†	139	130	6%	9
Understanding Growth (final section)	Transparency	139	129	7%	10

* Clear Articulation comprises four items administered at the beginning of the Learning Expectations section (Q15–Q19) and one additional item administered after Universal Expectations (Q31).

† Clear Calibration comprises two items administered in the Assessment section (Q36–Q37) and one additional item administered near the end of the Understanding Progression section (Q50).

Item Analysis

Item-level descriptive statistics showed that all response options were used across the five-point scale. As shown in Table E3, item means ranged from 2.66 to 4.31, and standard deviations ranged from 0.71 to 1.28, indicating adequate variability and no substantial floor or ceiling effects. Item nonresponse was generally low, ranging from 0% for items early in the survey to approximately 7% for items in the final section. This pattern is consistent with modest respondent attrition across the survey rather than construct-specific missingness.

Table E3. Summary of Teacher Survey Item Analysis Results

Construct	Items	Missing (%)	Mean Range	Corrected Item-Total Correlation
Clear Articulation	5	0.0–2.9	2.83–4.31	.40–.69
Developmental Progression	4	0.0–0.7	3.32–4.04	.59–.80
Learning Targets	4	0.7–2.2	3.25–4.04	.46–.70
Universal Expectations	3	2.2–2.9	3.24–3.66	.45–.62
Demonstration Location	3	3.6–4.3	2.66–3.43	.15–.39
Clear Calibration	3	4.3–5.0	3.20–3.32	.41–.66
Rigorous Application	3	5.0–5.8	3.37–3.93	.60–.80
Mastery-Based Progression	4	5.0–5.8	3.47–3.74	.63–.80
Iterative Purpose	3	5.0–6.5	3.62–3.88	.43–.62
Transparency	3	6.5–7.2	2.70–2.96	.80–.88

Corrected item-total correlations indicated generally strong internal coherence or discrimination across most CBL markers. Items within *Developmental Progression*, *Learning Targets*, *Rigorous Application*, *Mastery-Based Progression*, and *Transparency* demonstrated moderate to strong item-total correlations (approximately $r = .46$ to $.88$), indicating that individual items consistently contributed to their intended constructs. *Clear Articulation*, *Universal Expectations*, *Clear Calibration*, and *Iterative Purpose* also showed acceptable item discrimination overall. However, several items exhibited comparatively weaker relationships with their respective scales ($r \approx .40$ to $.51$), suggesting opportunities for future refinement rather than immediate concern.

Specifically, the *Demonstration Location* marker demonstrated the weakest psychometric performance. Corrected item-total correlations ranged from .15 to .39, which is substantially lower than those for the other constructs. Notably, the reverse-coded item assessing whether only school personnel can determine mastery had a corrected item-total correlation of .15, indicating limited consistency with the other two items. These results suggest that respondents may differentiate between who has the authority to determine mastery and whether learning experiences or evidence from outside of school contribute to mastery determinations.

This raises questions about whether these items represent a single latent construct. Further investigation through confirmatory factor analysis is warranted and may indicate the need to refine or reconceptualize the construct.

Internal Consistency

Overall, the item analysis provided preliminary support for the internal consistency of nine of the ten CBL markers, as shown in Table E4. Cronbach's alpha coefficients ranged from .71 to .92, excluding *Demonstration Location*. Most items functioned as intended and demonstrated sufficient variability and discrimination to warrant confirmatory factor analysis.

The *Demonstration Location* construct requires additional conceptual and psychometric refinement before being incorporated as a subconstruct in the CBL Measurement Framework or used in subsequent validity and research analyses. Until additional evidence supports its dimensionality and reliability, the individual items may be interpreted and reported separately as indicators of distinct aspects of learning beyond the classroom.

Table E4. Summary of Teacher Survey Reliability Coefficients for CBL Markers

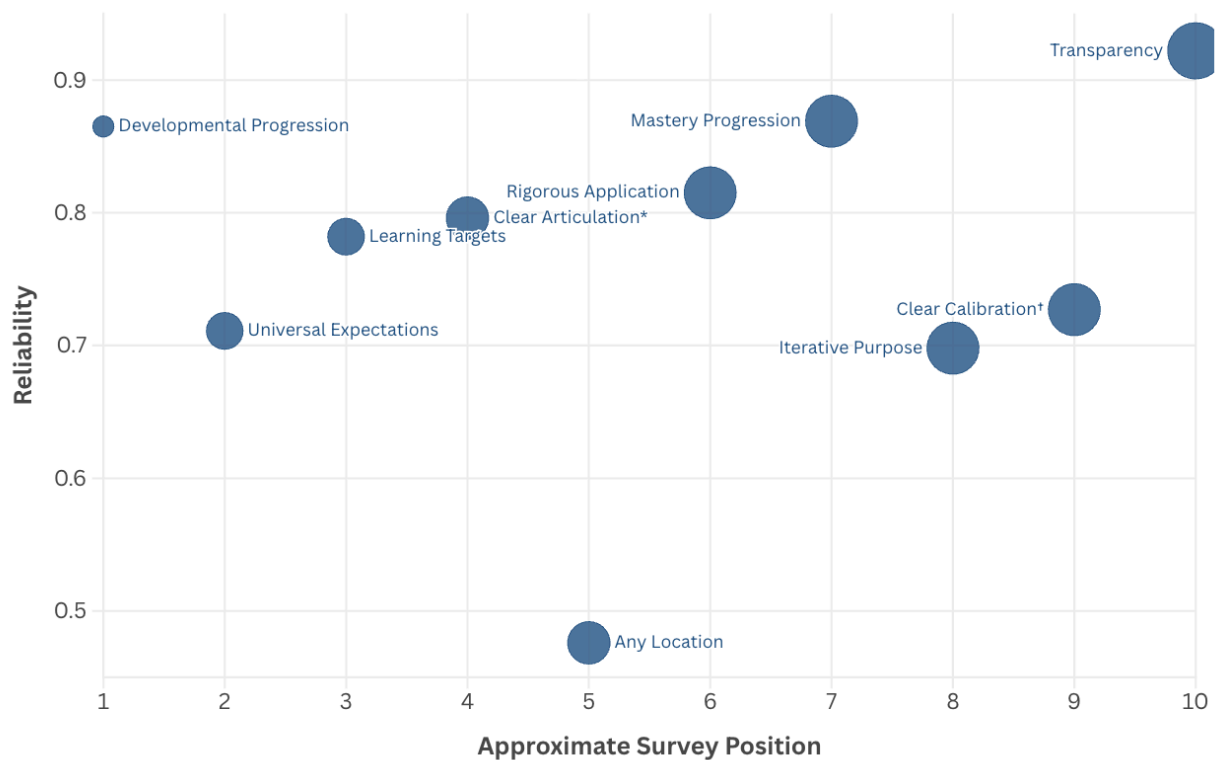
CBL Marker	Items	N	Reliability (α)
Clear Articulation	5	134	0.796
Developmental Progression	4	137	0.865
Learning Targets	4	135	0.782
Universal Expectations	3	135	0.711
Demonstration Location	3	133	0.476
Clear Calibration	3	130	0.727

Rigorous Application	3	131	0.815
Mastery-Based Progression	4	131	0.869
Iterative Purpose	3	130	0.698
Transparency	3	129	0.922

Note. Coefficient alpha values closer to 1 indicate greater internal consistency among the items within a scale. Lower coefficient alpha values indicate weaker internal consistency and suggest that the items may not function cohesively as indicators of the intended construct.

Although item nonresponse increased gradually across the survey, there was no evidence that internal consistency reliability systematically varied with survey position, as shown in the figure below. Markers administered near the end of the survey demonstrated reliability comparable to—or higher than—those administered earlier. For example, *Transparency*, the final marker, exhibited the highest internal consistency, whereas *Demonstration Location*, which appeared near the middle of the survey, demonstrated the lowest reliability. This pattern suggests that differences in reliability reflect the coherence of the proposed latent constructs rather than respondent fatigue or survey position.

Figure E1. Scatterplot of Reliability Coefficients and Relative Position of CBL Markers on Survey



Note: Bubble size represents the percentage of missing responses.

Correlations

Pearson correlation coefficients were calculated to assess the relationships among the proposed CBL markers and to provide preliminary evidence of convergent and discriminant validity. As anticipated, the CBL markers were generally positively correlated, with coefficients ranging from $r = .22$ to $.79$ after excluding the *Any Location* marker. This pattern indicates that the markers represent related dimensions within a broader competency-based learning framework while remaining empirically distinct.

The strongest associations were found among constructs related to students' understanding of learning expectations and progression.

- *Developmental Progression* exhibited strong positive correlations with *Learning Targets* ($r = .79$), *Universal Expectations* ($r = .67$), *Clear Articulation* ($r = .64$), and *Mastery-Based Progression* ($r = .64$).
- *Mastery-Based Progression* and *Iterative Purpose* were strongly correlated ($r = .73$), which aligns with the theoretical expectation that opportunities for revision, multiple assessment

attempts, and iterative feedback are essential mechanisms for student progression toward mastery.

Although these constructs are closely related, the correlation remained below commonly accepted thresholds for construct redundancy, indicating that they represent complementary rather than interchangeable dimensions of competency-based learning.

Moderate positive correlations were observed among most of the remaining constructs ($r = .33 - .60$), supporting the expectation that competency-based learning practices co-occur within schools while representing distinct instructional mechanisms. Notably, no correlations exceeded $.80$, which provides preliminary evidence that the proposed markers exhibit adequate discriminant validity.

In line with the item analysis, the *Any Location* construct exhibited substantially weaker relationships with the other CBL constructs ($r = .08-.38$). Correlations were particularly low with *Iterative Purpose* ($r = .08$), *Universal Expectations* ($r = .14$), and *Mastery Progression* ($r = .15$), indicating that teachers' perceptions of whether student learning can be demonstrated or evaluated outside the classroom may constitute a conceptually distinct dimension rather than an essential component of the broader competency-based learning framework. These findings suggest that the *Any Location* construct requires further refinement and validation before inclusion as a subconstruct within the CBL Measurement Framework.

Overall, the correlation analysis offers preliminary evidence that the proposed CBL measurement framework consists of a coherent set of related yet empirically distinguishable dimensions. The observed pattern of correlations is consistent with the hypothesized conceptual framework.

CBL Markers Construct-Level Descriptive Statistics

Descriptive statistics at the construct level were calculated to summarize educators' responses across the proposed competency-based learning (CBL) markers and to assess the distributional properties of each construct. Mean scores on the five-point response scale ranged from 2.85 to 3.78 (see Table E5). *Iterative Purpose* ($M = 3.78$, $SD = 0.67$), *Mastery Progression* ($M = 3.66$, $SD = 0.76$), and *Rigorous Application* ($M = 3.65$, $SD = 0.72$) had the highest average scores, while *Transparency* had the lowest ($M = 2.85$, $SD = 1.15$).

Table E5. Descriptive Statistics for CBL Markers

CBL Markers	n	mean	sd	median	trimmed	min	max	range	skew	kurtosis	se
Clear_Articulation	139	3.33	0.6455	3.40	3.35	1.00	4.80	3.80	-0.528	0.984	0.055
Developmental_Progression	139	3.56	0.7422	3.50	3.60	1.00	5.00	4.00	-0.739	1.235	0.063
Learning_Targets	138	3.55	0.6879	3.50	3.55	1.00	5.00	4.00	-0.528	1.733	0.059
Universal_Expectations	136	3.44	0.8160	3.67	3.47	1.00	5.00	4.00	-0.436	0.127	0.070
Clear_Calibration	134	3.26	0.7665	3.33	3.26	1.00	5.00	4.00	-0.228	0.082	0.066
Rigorous_Application	132	3.65	0.7193	3.67	3.70	1.00	5.00	4.00	-0.918	1.650	0.063
Mastery_Progression	132	3.66	0.7595	3.75	3.67	1.00	5.00	4.00	-0.315	0.155	0.066
Iterative_Purpose	132	3.78	0.6702	4.00	3.79	1.67	5.00	3.33	-0.466	0.204	0.058
Transparency	130	2.85	1.1473	3.00	2.83	1.00	5.00	4.00	0.166	-0.891	0.101

Note. Any Location was excluded from scale construction because the items did not demonstrate adequate internal coherence. Nevertheless, the construct was retained for subsequent analyses as a theoretically important component of competency-based learning warranting additional conceptual and psychometric refinement.

All constructs showed sufficient variability, with standard deviations between 0.65 and 1.15 and observed score ranges covering at least 3.33 points on the five-point scale. *Transparency* showed the greatest variability ($SD = 1.15$), while *Iterative Purpose* showed the least ($SD = 0.67$). No constructs demonstrated substantial floor or ceiling effects.

The distributions of the constructs were approximately normal. Skewness values ranged from -0.92 to 0.17, indicating generally symmetric distributions with only modest negative skew for some constructs. Kurtosis values ranged from -0.89 to 1.73, which are within commonly accepted thresholds for univariate normality. These results suggest that the construct scores have sufficient variability and appropriate distributional properties to support further correlational and latent variable analyses, given a sufficient sample size.

Although differences in construct means may indicate variation in educators' perceptions of competency-based learning practices, these descriptive statistics should be interpreted with caution. The primary aim of this study was to evaluate the psychometric properties of the CBL Measurement Framework, not to estimate the prevalence of competency-based learning practices. Therefore, construct means are reported to describe score distributions and assess the suitability of the measures for subsequent validity and research analyses alongside multiple study measures, rather than to draw substantive conclusions about implementation.

Fit of the Hypothesized Factor Structure

The hypothesized nine-factor CFA converged successfully and produced an admissible solution. There were no negative observed-item residual variances, no standardized loading point estimates greater than 1, and no latent factor correlations exceeding 1.

As shown in Table E6, the robust model chi-square was statistically significant, $\chi^2(428) = 891.18$, $p < .001$. Approximate fit indices also fell outside the prespecified reference values.

Table E6. Overall Fit of the Nine-Factor Teacher-Survey Model

Fit Statistic	Result
Robust χ^2	891.18
Degrees of freedom	428
p	< .001
CFI	.820
TLI	.792
RMSEA	.088

RMSEA 90% confidence interval	[.080, .096]
SRMR	.094

Note. CFA results use the expanded sample of 141 respondents. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual.

Although the model yielded an admissible solution with meaningful indicator loadings, the global fit statistics indicate that the specified nine-factor simple-structure model did not adequately reproduce the observed relationships among teacher-survey items. The CFA therefore should not be interpreted as fully confirming the hypothesized nine-factor structure.

Standardized Factor Loadings

All 32 indicators loaded positively and were statistically significant ($p < .001$) on their hypothesized factors. Standardized loading point estimates ranged from .480 to .968, and none fell below the prespecified .40 review threshold (Table E7).

Transparency showed the strongest and most consistent loadings, ranging from .831 to .934. *Developmental Progression*, *Mastery-Based Progression*, and *Rigorous Application of Learning* also showed generally strong loadings. Indicator strength was more variable for *Clear Articulation*, *Connection to Learning Targets*, *Clear Calibration*, and *Iterative Purpose*.

Table E7. Standardized Factor Loadings for the Nine-Factor Teacher-Survey Model

CBL Marker	Indicator Description	Standardized Loading	95% CI
Clear Articulation	Students use competencies to plan work across classes	.783	[.648, .918]
	Students use competencies to plan learning outside classes	.791	[.662, .920]
	Students identify competencies across classes	.587	[.371, .803]
	Students identify competencies in other parts of life	.668	[.480, .857]
	Competencies prepare students for their future	.480	[.212, .749]
Developmental Progression	Choose experiences supporting higher competency levels	.627	[.467, .788]

	Students explain improvement at higher levels	.761	[.642, .880]
	Students explain the competency level they are working toward	.866	[.816, .917]
	Students explain how close they are to mastery	.906	[.861, .950]
Connection to Learning Targets	Students explain concepts and skills being worked on	.637	[.471, .802]
	Students explain closeness to mastery	.848	[.771, .926]
	Students explain requirements for each level	.797	[.729, .865]
	Students use a blueprint/rubric	.537	[.377, .696]
Universal Application of Expectations	Use rubric describing mastery criteria	.647	[.420, .873]
	Same mastery criteria across learning experiences	.776	[.632, .919]
	Same mastery criteria expected across adults/settings	.611	[.398, .823]
Clear Calibration	Results consistent regardless of assessor	.809	[.692, .927]
	Results consistent regardless of assessment method	.834	[.722, .947]
	Use assessment results to determine next learning steps	.541	[.315, .767]
Rigorous Application of Learning	Apply learning to authentic/community contexts	.705	[.557, .853]
	Use meaningful performance tasks	.683	[.557, .809]
	Apply multiple skills to authentic problems	.968	[.866, 1.070]
Mastery-Based Progression	Continue working until mastery	.777	[.685, .870]
	Move on when ready	.664	[.554, .774]
	Continue working until criteria are met	.891	[.844, .939]
	Improve work until mastery criteria are met	.854	[.788, .920]

Iterative Purpose	Revise or practice before final credit	.761	[.641, .881]
	Retake assessments	.551	[.370, .731]
	Use multiple drafts or attempts	.693	[.542, .845]
Transparency	Maintain tracker/portfolio organized by competencies	.914	[.864, .964]
	Use tracker/portfolio to monitor progress	.934	[.883, .985]
	Share tracker/portfolio for feedback	.831	[.740, .923]

Note. Indicator descriptions are abbreviated analytic labels rather than the full administered survey wording. All standardized loading point estimates were statistically significant at $p < .001$. The confidence interval for the *Rigorous Application of Learning* item concerning multiple skills extended above 1, reflecting imprecision in the estimate; the standardized point estimate remained below 1.

The weakest loading was the *Clear Articulation* item assessing whether competencies prepare learners for their future, $\lambda = .480$, 95% CI [.212, .749]. Although statistically significant, this estimate was weaker and less precise than most other indicators.

Local Model Diagnostics

Standardized residuals and modification indices identified several localized sources of model misfit.

The largest standardized residual involved the *Connection to Learning Targets* item referencing a blueprint or rubric and the *Universal Application of Expectations* item referencing a rubric describing mastery criteria (residual = .392). Because both explicitly reference rubrics and mastery criteria, educators may interpret them as reflecting overlapping aspects of the same learner experience or instructional practice.

A second pattern involved the *Clear Calibration* item assessing use of assessment results to determine next learning steps. The item had a moderate primary loading on *Clear Calibration* ($\lambda = .541$) but also showed relationships with indicators associated with *Iterative Purpose*, *Mastery-Based Progression*, *Developmental Progression*, and *Connection to Learning Targets*. Educators may therefore experience assessment-informed next steps as spanning several aspects of CBL rather than as uniquely associated with calibration.

The *Clear Articulation* item concerning preparation for learners' futures also had the weakest primary loading and showed relationships with indicators from other constructs, suggesting that it may reflect broader perceived relevance of competencies.

These diagnostics identify priorities for future item refinement. They were used to interpret the measurement evidence rather than to modify the model. No cross-loadings, residual covariances, or other post hoc parameters were added to the reported CFA.

Latent Factor Relationships and Discriminant Validity

All nine latent factors, as shown in Table E8, were positively correlated, with correlations ranging from .202 to .909. The strongest relationship was between *Mastery-Based Progression* and *Iterative Purpose* ($r = .909$), followed by *Developmental Progression* and *Connection to Learning Targets* ($r = .887$).

These latent-factor estimates differ from the observed marker-score correlations described earlier because they estimate relationships among the hypothesized constructs within the CFA while accounting for measurement error.

The magnitude of the two highest latent correlations prompted targeted follow-up analyses comparing freely estimated models with models constraining each selected correlation to 1.00.

Table E8. Targeted Discriminant Validity Analyses

Factor pair	Correlation in Primary CFA	Scaled $\Delta\chi^2$	df	p	Interpretation
Developmental Progression and Connection to Learning Targets	.887	18.67	1	< .001	Constraining the correlation to 1.00 significantly worsened fit, supporting empirical distinction in this comparison.
Mastery-Based Progression and Iterative Purpose	.909	2.82	1	.093	Constraining the correlation to 1.00 did not significantly worsen fit; empirical distinction was not established.

Note. Correlations are estimated in the primary nine-factor CFA. Difference-test results are targeted Satorra-Bentler scaled chi-square comparisons using the expanded teacher sample.

The analysis supported empirical distinction between *Developmental Progression* and *Connection to Learning Targets* despite their strong association.

In contrast, the teacher data did not establish discriminant validity between *Mastery-Based Progression* and *Iterative Purpose*. Constraining their correlation to 1.00 did not significantly worsen model fit. This finding should not be interpreted as demonstrating that the two constructs are conceptually identical. Their high latent correlation and statistical indistinguishability in this sample nevertheless indicate that their measurement boundaries require additional examination.

These targeted analyses address two highly correlated factor pairs. They do not resolve the primary model's overall misfit or establish that all nine factors are clearly distinguishable.

Interpretation and Analytic Limitations

The teacher-survey analyses provide mixed but informative evidence for marker-level interpretation.

The preliminary analyses showed generally coherent item sets for nine of the ten proposed marker scales. Those nine scales had internal consistency coefficients ranging from approximately .70 to .92 and meaningful variation in observed scores. The CFA similarly found that all 32 included indicators loaded positively and significantly on their hypothesized factors.

At the same time, the global CFA fit indices did not support full confirmation of the exact nine-factor simple-structure model. Local diagnostics identified several items whose content appeared to span marker boundaries, and the targeted discriminant-validity analyses did not establish empirical distinction between *Mastery-Based Progression* and *Iterative Purpose*. The evidence therefore supports qualified marker-level interpretation rather than strong claims that all proposed factors operate as fully independent dimensions.

The current *Demonstration Location* items present a separate measurement limitation. Their weak item-total correlations and internal consistency do not support treatment as a single marker-level composite. This finding concerns the current survey operationalization rather than the conceptual status of *Demonstration Location* within the broader marker framework.

Interpretation is also bounded by the modest teacher sample relative to the complexity of the 32-indicator, nine-factor CFA. Some estimates had comparatively wide confidence intervals, sparsely populated response categories may contribute additional uncertainty, and the 141 respondents represented nine participating high schools. The analysis does not establish that the measurement structure operates equivalently across high schools or would replicate in an independent teacher sample.

Together, the findings support using teacher responses as **descriptive evidence about educators' perceptions of learner experiences and related instructional practices associated with individual CBL markers**, with the qualifications described above. They do not support creation of a single overall teacher CBL score, treatment of Demonstration Location as a reliable three-item composite, or strong claims that all nine modeled dimensions are empirically independent.

Future research can refine the identified items, test revised or alternative measurement structures, and examine the constructs in larger and independent teacher samples. Within this study, teacher-survey findings are interpreted as one source of evidence within the broader measurement framework rather than as a stand-alone representation of learners' CBL experiences.

Appendix F.

Learner Focus Group Methods and Analysis

Purpose

Learner focus groups examined how students described their experiences of learning within the ten participating high schools. They provided opportunities to discuss experiences extending beyond a single classroom visit, including how learners understood competencies, recognized progress, used feedback, and demonstrated learning across school and other settings. Focus groups also allowed participants to elaborate on and respond to one another’s accounts.

Focus-group analysis contributed primarily to RQ4 and RQ5 by examining which of the ten hypothesized CBL markers appeared in learners’ accounts and how the breadth of those experiences varied within scorable evidence. Learner accounts also provided contextual and explanatory evidence that supported interpretation of findings from other study measures.

- RQ4. Which of the ten hypothesized markers of competency-based learning were evident within and across participating high schools and evidence sources?
- RQ5. How did learners and educators perceive and describe these markers, and what variation in breadth was visible in scorable classroom-observation and learner focus-group evidence?

Participants and Data Collection

Researchers conducted 21 learner focus groups with 105 participants across the ten participating high schools during spring 2026 site visits. Bismarck Public Schools’ Empower[Ed] participated in three focus groups; each of the other nine high schools participated in two. Recruitment procedures are described in *Appendix A*.

Table F1 summarizes participation in focus groups across the ten high schools.

Table F1. Learner Focus Group Participation

High school	Focus Groups	Learner Participants
Bismarck Public Schools’ Empower[Ed]	3	13
Clark Street Community School	2	12

Compass Community Collaborative School	2	9
Crosstown High School	2	10
Da Vinci Communications	2	14
Del Lago Academy	2	13
Harwood Union High School	2	8
Lindsay High School	2	9
Northern Cass	2	9
STEAM Academy	2	8
Total	21	105

The semi-structured protocol was designed for approximately 45 minutes. Questions began with learners' daily experiences and aspirations and then explored how they understood competencies, developed proficiency, received and used feedback, and demonstrated learning. Follow-up prompts invited concrete examples and clarification. The protocol used accessible language so that learners could describe their experiences using terminology familiar within their high school rather than needing to recognize the study's marker names.

Focus groups were audio recorded with participants' permission and transcribed for analysis. Participant permission, assent, and consent procedures are described in *Appendix A*. The substantive focus-group questions are reproduced at the end of this appendix; assent language, recording instructions, facilitator transitions, and other operational directions have been omitted from the published version.

Coding and Analysis

All transcripts were reviewed, cleaned, and de-identified prior to coding. Names and other directly identifying information were removed or replaced with neutral redaction labels while preserving the substantive content of learners' accounts.

Identifying Evidence of the Markers

Researchers analyzed learner accounts using the marker codebook and hypothesized breadth-of-experience continuum described in *Appendix B*. Coding focused on excerpts that contained evidence relevant to one or more of the ten hypothesized markers. The excerpt served

as the analytic unit rather than the individual learner, focus group, or high school. Using explicit coding categories and defined units of verbal evidence is consistent with methodological guidance for systematic analysis of verbal data in learning research (Chi, 1997).

The analysis was primarily deductive: it began with the study's ten hypothesized markers and examined evidence relevant to those constructs rather than attempting to generate a comprehensive account of every aspect of high school experience discussed by participants. The topic headings in the focus-group protocol guided discussion but did not determine how learners' responses were coded. For example, an excerpt arising during a discussion of assessment could provide evidence related to feedback and revision, application of competencies, progression, or learners' understanding of expectations.

Coding depended on the substance of the learner's account and the evidence available to connect that account to a marker. An engaging project, flexible learning experience, or activity outside the high school did not by itself constitute evidence of a particular CBL marker. The excerpt or context surrounding the excerpt needed to contain enough information to establish the relevant connection to competencies and the learner experience represented by the marker.

Interpreting Breadth of Experience

When an excerpt contained sufficient evidence, researchers interpreted it using the four levels of the hypothesized CBL continuum described in *Appendix B: Pre-Emergent, Baseline, Maturing, and Advanced*. The continuum provided a common language for examining variation in the role of competencies, learner ownership, and the scope of an experience. Evidence that did not contain enough information to support a continuum interpretation was not assigned a breadth level.

Continuum indicators described what was evident within a particular excerpt. They did not establish a learner's typical experience across courses, settings, or time and were not used to evaluate the overall practice of a high school. The same learner could describe experiences reflecting different levels of breadth depending on the competency, task, course, or setting being discussed.

Researcher Review

Two researchers participated in coding and review. One researcher completed the initial coding of the transcripts using the marker codebook and hypothesized breadth-of-experience continuum. A second researcher subsequently conducted a secondary review of the coding, providing an additional examination of marker assignments and breadth interpretations.

Because this process involved initial coding followed by secondary review rather than two researchers independently coding the same material in parallel, it was not treated as an independent double-coding design and was not used to calculate inter-rater reliability.

Summarizing Scorable Marker Evidence

Researchers summarized how frequently the coded focus-group evidence contained enough information to interpret each marker. Excerpts assigned to any level of the hypothesized continuum, from *Pre-Emergent* through *Advanced*, were counted as scorable evidence relevant to that marker. Focus-group rankings therefore reflected how frequently markers could be interpreted within the coded excerpts, not how frequently they received higher breadth classifications.

Frequency summaries provided a way to identify patterns across a substantial body of verbal evidence while retaining the underlying excerpts for qualitative interpretation. The use of numerical summaries alongside qualitative examination is consistent with Chi's (1997) approach to quantifying coded verbal data.

These frequencies describe the coded excerpts, not the prevalence of an experience among participating learners. A learner could contribute multiple scorable excerpts, and the amount of evidence associated with a marker could be affected by the detail of learners' accounts, follow-up questions, and the direction of a group's conversation. Frequency results therefore should not be interpreted as the percentage of learners who experienced a marker or as estimates of how often the experience occurred in daily high school life.

Integrating Focus-Group Evidence

Focus-group evidence was interpreted alongside survey, classroom-observation, and high school contextual evidence. Learner accounts were particularly useful for explaining experiences that extended beyond the temporal and physical boundaries of a brief classroom observation, including learning that developed across courses, settings, educators, or longer periods.

Integration did not require different evidence sources to produce identical findings. Convergence, complementarity, inconsistency, and divergence can each contribute to interpretation in mixed-methods research (Mathison, 1988). In this study, focus groups provided a learner-centered account of how competency-based experiences were understood and connected across contexts rather than functioning as a standardized measure of prevalence.

Interpretation and Limitations

Focus groups provided detailed accounts of learning across settings and over time, although the conversations themselves occurred during a single spring 2026 data-collection window. They therefore expanded the temporal context available through brief classroom observations without providing longitudinal evidence of how individual learners' experiences changed over time.

The findings reflect the accounts of participating learners rather than a representative sample of all learners within participating high schools. Learners were purposefully recruited with support from high school staff, as described in *Appendix A*.

Group interaction can prompt participants to elaborate, recall experiences raised by peers, or consider perspectives other than their own, but it can also influence which participants speak, what they are willing to share, and which perspectives become prominent (Wilson, 1997). Recall, speaking opportunities, group dynamics, and facilitator follow-up also shaped the evidence available for analysis.

The focus-group protocol did not provide equal opportunities for every marker to surface and prioritized markers hypothesized to be more challenging to observe in a non-interactive classroom visit. Some questions addressed particular experiences directly, and some prompts introduced examples of practices before asking learners whether they had encountered them. In addition, beginning the analysis with predefined markers directed analytic attention toward the experiences hypothesized by the study. Findings should therefore be interpreted as an examination of those hypotheses rather than an exhaustive description of learner experience.

The absence of scorable evidence does not establish that a marker was absent from a learner's experience or from a high school. Conversely, a detailed excerpt establishes that a learner described an experience; it does not establish that the experience was shared by all learners at that high school. Breadth classifications were therefore applied to excerpts and were not aggregated into high school-level evaluations.

Within these boundaries, the focus groups provide contextual and explanatory evidence about how learners understand and describe competency-based learning across different high school models. Their accounts help make visible experiences that may unfold across courses, settings, and time while providing evidence that can be interpreted alongside the study's other measures.

Learner Focus Group Questions

The questions below preserve the substantive content of the study protocol. Assent language, recording instructions, facilitator transitions, and closing language have been removed.

Opening Question

1. Please share your first name, the pronouns you use, and one thing you like to do in your free time.

School Competencies

2. When you think about a typical day, what kinds of things are you learning? Can you walk me through what that looks like?

3. When you think about your time in high school overall, what do you hope to be really good at by the time you graduate?

Follow-up:

- Are these the same kinds of things your school wants you to be really good at, too?

Mastery-Based Progression

4. How clear is it to you what you need to do to improve or get better at the things you've just named?

Follow-up questions:

- Does your school give you any kind of roadmap or guide that shows how skills like the ones you mentioned can develop over time?
 - Think about a time when you were trying to improve a skill or better understand something. How did you know what to focus on and how to improve?
5. In some schools, students move on to the next topic only after they've shown they've really learned it. That can mean students are working on different things at the same time; some might move ahead more quickly, while others take more time to get there. Is that something you've experienced in your classes?

Follow-up questions if learners describe this experience:

- How does it feel when you're at a different place in your learning journey than others?
- How do students stay on track so that no one is left behind by the end of the year?
- What helps you know when you're ready to move on?

Assessment: Demonstration Location and Rigorous Application of Learning

6. What classes have meant the most to you this year and why?

Follow-up questions:

- Can you tell me more about what assessment looks like in these classes or outside experiences?
 - How have you learned about yourself from the feedback you've received from those assessments and from folks evaluating your work?
 - Do you ever get to revise your work or retake an assessment based on feedback or assessment results?
7. What kinds of learning experiences have you had outside of school this year? What did you want to accomplish?

Follow-up questions:

- Were these experiences all for class credit?
- Did you do anything just for fun to get better at a skill, learn more about something, or explore an interest like entrepreneurship? If so, did those experiences count for class credit?

- What kind of evidence did you have to show that demonstrated you'd leveled up on a concept or skill you were working on at school?
 - Who reviewed and verified that evidence?
- 8. If learning is assessed outside of school, how similar or different is assessment inside versus outside of school?
- 9. Do you feel like your out-of-school learning is valued the same way as in-school work?

Closing Question

- 10. What's something you want others to know about your school?

Appendix G.

Policy Focus Groups, Interviews, and Documentary Analysis

Purpose and Scope

Policy focus groups and interviews examined how state and local policy conditions surrounded and influenced the implementation of competency-based learning (CBL) across participating high schools. The inquiry considered how leaders understood available authority, translated policy into local practice, navigated constraints, and accessed implementation support.

This appendix describes the policy protocols, documentary evidence, and cross-site analytic approach. *Appendix A* provides the overall study design, sample counts, and participant protections. The companion policy report presents the substantive policy findings.

Protocols and Facilitation

Two semi-structured protocols guided data collection: the *Policy Leader Focus Group* protocol and the *Policy Leader Interview* protocol. Both addressed understandings of CBL, enabling and constraining policy conditions, variation across contexts, implementation responses, and opportunities for closer alignment between policy and practice.

The focus-group protocol emphasized comparison across participants' contexts. Questions moved from shared understandings of CBL to examples of enabling policies, implementation challenges, local responses, and future priorities. Follow-up prompts explored what made policies usable, where high schools encountered friction, and what tradeoffs accompanied efforts to work within existing requirements. Sessions were designed for approximately 60 minutes and established expectations for respectful discussion, participation, and confidentiality.

The interview protocol addressed the same broad domains while allowing deeper exploration of the participant's role, system, and policy context. Questions examined how CBL was understood locally, which policies supported implementation, where policy and practice diverged, and how high schools responded.

The distinction between formats matters for interpretation. Focus groups allowed participants to respond to one another, compare experiences, and surface differences across contexts. Contributions made within a group discussion should therefore not be treated as independent survey responses or assumed to represent agreement among all participants. Research on focus-group methodology emphasizes the importance of attending to these interactions when interpreting group evidence (Parker & Tritter, 2006).

Document Evidence

Policy conversations were interpreted alongside publicly available documents relevant to the ten participating high schools and their state or local policy contexts, FullScale and KnowledgeWorks' [State of CBE](#) policy map, and materials identified through the broader study landscape scan.

Relevant sources included state statutes and regulations, administrative guidance, state and local board policies, charter or authorization arrangements, credit and graduation requirements, accountability provisions, and high school or system procedures. Documents were included when they provided direct evidence about formal authority, operating requirements, guidance, or procedures relevant to a participating high school or its policy context.

Document and conversational evidence served different purposes. Documents established what was formally authorized, required, or described. Participants provided evidence about how those provisions were understood, interpreted, and used in practice, including implementation challenges that might not be apparent from written policy alone. Document analysis can therefore provide institutional context and an additional source against which participant accounts can be considered (Bowen, 2009).

A central distinction throughout the analysis was between the availability of a policy provision and the documented use of that provision by a participating high school. The existence of state-level flexibility was not treated as evidence that every high school in that state relied on it or used it in the same way.

Analysis and Cross-Site Interpretation

The policy analysis was designed as a complementary, exploratory component of the broader study rather than as a parallel measurement study. Its purpose was to identify and describe policy conditions surrounding implementation, not to generate standardized scores or formally test a predefined policy framework.

Analysis used a structured thematic approach. Broad analytic domains were defined in advance from the study questions and protocol topics, while more specific themes and examples within those domains were identified through review of the conversation content. One researcher conducted the primary review and coding of the policy conversations and developed within- and cross-conversation summaries.

Because the policy analysis did not use independent parallel coding or a formal secondary coding process, inter-rater reliability was not estimated. Findings were therefore interpreted as descriptive thematic evidence and were bounded accordingly.

Within each context, the analysis considered relationships among available authority, local governance, operating procedures, documentary evidence, and participants' accounts of implementation. Across conversations, the analysis examined recurring patterns and differences in how these conditions connected to CBL practice.

The synthesis was organized around four broad areas:

- policy pathways through which participating high schools operated;
- interactions among assessment, credit, staffing, accountability, and reporting requirements;
- implementation supports available to high schools and systems; and
- relationships through which local experience informed broader policy learning.

These areas organized interpretation of the policy evidence and were separate from the ten learner-experience markers and the hypothesized breadth-of-experience continuum used elsewhere in the study.

Differences among contexts were retained as part of the analysis rather than reduced to a single preferred policy model. For example, a formal authorization arrangement might establish operating flexibility, while local guidance and participant accounts could show how that flexibility was translated into assessment practices, credit decisions, or reporting procedures.

Conversation-Level Summaries

Where frequencies were reported, the conversation served as the unit of analysis. Each interview or focus-group session counted as one conversation. Counts indicated whether an issue surfaced within a conversation rather than how often it was mentioned or how many participants endorsed it.

The combined denominator was 17 conversations. Ten were site-focused conversations, and seven involved broader policy-system actors, including state agencies, nonprofit organizations, and intermediaries. These categories describe the primary focus of the conversations rather than their division into interviews and focus groups.

A topic mentioned multiple times within one conversation contributed one occurrence to a conversation-level count. These frequencies therefore describe the coverage of an issue within the available evidence. They do not measure participant consensus, the severity of a constraint, the strength of an enabling condition, or the prevalence of an issue across high schools more generally.

Issues for which a consistent cross-conversation count was not established were presented through illustrative evidence rather than assigned a frequency.

Connecting Policy and Learner-Experience Evidence

Policy evidence was considered alongside learner-experience evidence when the available sources supported a plausible connection.

For example, learner accounts of reassessment could be considered alongside local grading and credit procedures. Evidence developed through a community or work-based experience could be interpreted alongside policies governing who was authorized to assess that learning, what evidence could count, and how demonstrated learning could be recognized for credit.

These connections provided context for understanding the conditions surrounding learner experiences. They were not treated as tests of whether particular policies produced those experiences.

Policy conversations did not contribute to the ranking of CBL markers, receive breadth-continuum scores, or enter the quantitative analyses used to examine the learner-experience measures. The policy and learner-experience analyses retained distinct research questions, evidence sources, and units of analysis while contributing complementary perspectives on implementation.

Trustworthiness and Analytic Limitations

Use of common protocol domains supported comparison across policy conversations, while document evidence provided an additional source for contextualizing participants' accounts. Convergence across conversational and documentary sources increased confidence in an interpretation, while discrepancies were retained and examined rather than treated as errors requiring resolution.

At the same time, the policy analysis was less formalized than the study's primary learner-experience analyses. One researcher conducted the thematic review and coding, and the analysis did not include independent double coding, inter-rater reliability estimation, or a separate psychometric or measurement-validation process. The resulting findings should therefore be interpreted as descriptive and exploratory.

The evidence is also bounded by the participating high schools, organizations, and individuals. Leaders' accounts reflect their particular roles, responsibilities, and experiences. In focus groups, participation patterns and professional relationships may influence which perspectives are voiced and how openly participants respond. A statement made during a group discussion therefore does not establish agreement among everyone present (Parker & Tritter, 2006).

Public documents establish what is formally authorized, required, or described but do not demonstrate how consistently a policy or practice is used. Likewise, the absence of an issue from a conversation does not establish that it was absent or unimportant in that context.

Conversation-level frequencies describe how widely an issue surfaced within the available evidence; they should not be interpreted as measures of prevalence, severity, effectiveness, or consensus.

The analysis supports descriptive conclusions about relationships among policy conditions, local interpretation, and implementation within the settings represented in the study. It does not identify a necessary or sufficient combination of policies, explain differences in learner-marker scores, or attribute learner experiences or outcomes to particular policy arrangements.

Within these boundaries, the policy analysis provides a complementary view of the conditions surrounding CBL implementation and identifies areas where additional research and policy learning may clarify how formal authority, local interpretation, and implementation interact across contexts.

Appendix H.

Competency and Academic Outcome Methods

Purpose and Scope

The competency and academic outcome analyses examined what available records showed about learner progress and performance within the ten participating high schools. The analyses served two complementary purposes: describing outcomes relevant to each high school's learning model and, where compatible learner-level records were available, examining relationships between local competency measures and conventional academic indicators.

The analyses drew on high school-provided internal records and publicly reported data. Most outcomes describe the 2024–25 school year, although reporting periods, assessment windows, and graduating cohorts varied by measure and are identified alongside individual results. These outcomes precede the learner-experience data collected primarily in spring 2026 and therefore provide descriptive context rather than temporally aligned outcomes linked to the survey, observation, or focus-group measures.

Measures differed across participating high schools because local competency systems, data structures, and outcome priorities differed. The study did not create a common cross-site competency outcome or assume that similarly named or numbered measures represented equivalent constructs across high schools.

The analyses are descriptive and correlational. They do not estimate the causal effects of competency-based learning (CBL) or test whether learners' survey-reported CBL experiences produced academic outcomes. *Appendix A* describes the overall study design and collaborative outcome-blueprinting process; this appendix documents the analytic principles used to prepare and interpret the resulting outcome evidence.

Selecting Measures for Analysis

Building on the collaborative blueprinting process described in *Appendix A*, researchers identified records that could provide evidence about outcomes relevant to each high school's learning model. Potential measures included local competency ratings, mastery credits, performance demonstrations, course progress, standardized assessments, graduation indicators, advanced coursework, work-based learning, and postsecondary outcomes.

Site-specific data-preparation recorded available data sources, measure definitions, reporting periods, linkage requirements, limitations, and analytic decisions. Final analyses relied on the most complete verified records available for each measure. Measures considered during blueprinting

were not reported when records were unavailable, insufficiently documented, outside the relevant reporting period, or could not be linked or interpreted with adequate confidence.

The absence of a measure from the analytic dataset does not establish that a high school lacked the corresponding practice or information. It indicates that the available records did not support the intended analysis.

Establishing Analytic Populations and Preparing Records

Defining Populations and Denominators

Analytic populations were defined separately for each measure using the available enrollment records, high school identifiers, grade levels, reporting periods, and applicable participation requirements. A single master dataset was therefore not used automatically as the denominator for every outcome.

Enrollment counts, tested populations, course participants, graduates, and learners with linked competency records represent different analytic populations. The denominator reported for each result reflects the population to which that measure applies and the usable records available for analysis.

Reporting periods were also verified for each source. Records were included in 2024–25 summaries only when their time period and population supported that interpretation. Measures covering other reporting periods, namely publicly available datasets, were either identified explicitly with those periods or excluded from 2024–25 outcome summaries.

Where applicable, records outside the authorized population or reporting window were excluded. Documented opt-out and eligibility restrictions were also incorporated into site-specific preparation procedures.

Linking Records

Learner-level records were linked using shared identifiers when they were available. When direct identifiers were unavailable across source files, researchers used documented site-specific matching procedures based on the information available in those records. Ambiguous or unmatched records were not treated as successful matches. Outcome data was not linked to student survey records as the school years and significant portions of the sample were not consistent.

Analyses requiring information from two sources were restricted to learners with usable linked information on both measures. As a result, analytic samples for correlations and other linked analyses could be smaller than the total number of learners represented in either source independently.

Organizing Different Units of Evidence

Source files represented different units of information, including learners, courses, assessment administrations, competency ratings, examination records, and participation entries. Researchers retained these distinctions when preparing analytic variables.

For example, the number of AP examinations represented in a dataset is different from the number of learners who took at least one AP examination. Similarly, participation in a course is conceptually different from successful completion of that course or the number of credits earned.

Researcher-derived variables were documented separately from high school-provided summary fields. Derived measures were interpreted according to the records and decision rules used to construct them rather than presented as official high school measures unless the source itself defined them that way.

Missing, Unmatched, and Nonnumeric Records

Missingness was interpreted in relation to the source and measure. Unless source documentation established that the absence of a record represented zero attainment or nonparticipation, missing and unmatched records were treated as unknown rather than coded as zero.

A learner without a competency record, for example, could have incomplete documentation rather than no competency attainment. Similarly, the absence of a participation entry could reflect incomplete tracking rather than proof that an experience did not occur.

Nonnumeric status codes were also kept distinct from numeric scores. Statuses such as in-progress or incomplete were interpreted according to their documented meaning and were not automatically treated as failing grades, zero proficiency, or numeric credit deficits.

These distinctions governed both calculations and reporting language. Results were labeled according to what the available records established, with qualifications where coverage or documentation was incomplete.

Defining Outcome Measures

Local Competency Evidence

Local competency measures retained their high school-specific meanings. The study did not combine these measures into a common cross-site score or assume that similar numeric scales reflected equivalent definitions of proficiency.

Available competency evidence varied in content, unit, and purpose. Some sources represented proficiency judgments, others documented mastery credits, and others summarized performance across courses or demonstrations. Table G1 illustrates several forms of competency evidence represented in the study and the interpretive decisions associated with them.

Table H1. Examples of Local Competency Evidence and Interpretation

High School	Available Evidence	Interpretation Considerations
Bismarck Public Schools' Empower[Ed]	Standards mastery and Career Readiness Practices measures	The measures represent distinct dimensions of learning and retain their own definitions and reporting periods.
Clark Street Community School	Foundational and advanced mastery credits documented in graduate transcripts	Foundational requirements are common to graduates and may provide limited variation within that population. Advanced credits provide a different view of demonstrated learning. Graduate transcripts do not represent all currently enrolled learners.
Compass Community Collaborative School	Competency ratings reported across quarters	Analytic summaries reflected the available rating records according to the documented site-specific derivation. Learners without records were not assumed to have zero attainment.
Harwood Union High School	Learning Expectation and course scores	These are distinct measures. Nonnumeric statuses were excluded from numeric averages rather than assigned a score of zero.
Northern Cass High School	Course-level standards proficiency records	Scores represent proficiency within local courses and were interpreted in relation to the competencies represented in those courses and the academic measures used in each analysis.

STEAM Academy	Senior Defense of Learning ratings for STEAM Habits and presentation performance	The records describe a senior demonstration and its assessed dimensions rather than repeated measures of competency development across all grade levels.
---------------	--	--

Note. The examples illustrate forms of local evidence available for analysis. They do not imply that every available variable was reported. Site profiles identify the specific measures used in published findings.

Academic, Participation, and Transition Measures

Conventional academic evidence included available course progress, GPA, standardized assessments, graduation indicators, and postsecondary records. Participation measures included opportunities such as advanced coursework, career and technical education, dual enrollment, and documented learning beyond conventional coursework.

Analyses distinguished participation from completion and intentions from verified outcomes. Enrollment in a course did not establish successful completion or credits earned. A postsecondary plan did not establish subsequent enrollment. A list of graduates did not provide the cohort denominator required to calculate a graduation rate.

Below, Table H2 summarizes common calculation and denominator rules applied across analyses.

Table H2. Measure Definitions and Denominator Considerations

Measure	Calculation or Analytic Unit	Denominator and Interpretation
Credit completion	Credits earned relative to available or attempted credits, as defined in the source	The denominator reflects credit opportunities represented in the records. A mean of learner-level completion percentages is distinct from a pooled ratio of all credits earned to all credits available; the reported calculation is identified with the result.
In-progress or incomplete coursework	Number of qualifying course records per learner	Results describe the documented course status and are not automatically equivalent to failed courses or numeric credit deficits.
AP exam participation	Unique learners with at least one qualifying examination record	Examination records establish participation among represented test takers. A high school-wide percentage requires an appropriate eligible learner denominator.

AP exam performance	Qualifying examination scores meeting the reported threshold	Examination-based percentages use examinations as the denominator. Learner-based percentages use learners and require a stated rule, such as earning at least one score of 3 or higher.
Standardized assessment performance	Available scores or performance categories for the specified subject and administration	Results represent eligible tested learners with usable records, who may differ from total high school enrollment.
CTE or dual-enrollment participation	Learners with qualifying course records	Course participation is distinct from pathway completion, successful course completion, or credits earned.
Documented real-world learning	Learners with qualifying entries in available tracking records	Results describe documented experiences. Incomplete tracking does not support interpretation as a comprehensive participation rate.
Postsecondary transitions	Applications, reported plans, or verified enrollment, depending on the source	Each outcome retains its own definition and population. Immediate enrollment does not establish persistence or completion.

Note. Exact reporting periods, populations, thresholds, and sample sizes accompany results in the site profiles. Publicly reported measures retain the definitions used by the corresponding reporting agency.

Describing Learner Outcomes

Counts and percentages were used to summarize categorical outcomes and participation. Means, medians, and distributional summaries were used for numeric measures where appropriate. Sample sizes reflected the usable population for each measure rather than one common analytic sample across all outcomes.

Assessment results were interpreted according to their corresponding reference systems. National percentile scores describe performance relative to an assessment's norm group, while state performance categories describe attainment relative to state-defined expectations. These metrics were not treated as interchangeable.

Achievement and growth were also distinguished. Where growth was analyzed or reported, interpretation depended on the relevant assessment windows and availability of matched records. The existence of records from multiple time points was not itself treated as evidence of learner growth. Most outcomes reported in the site profiles represent descriptive snapshots rather than longitudinal estimates of change.

Selecting and Interpreting Contextual Comparisons

Contextual reference points were used to help readers interpret learner outcomes relative to relevant external benchmarks. Depending on the measure and available public reporting, reference points could include another local high school, a district, a county, or statewide results.

Selection considered geographic and organizational relevance, grade configuration, available demographic context, and availability of the same or a sufficiently comparable outcome measure and reporting period. State results were used when they provided an appropriate broader reference point.

These reference populations were not statistically matched comparison groups. Similarity on selected characteristics does not establish equivalence in prior achievement, enrollment processes, program offerings, learner needs, or other factors related to outcomes.

Researchers also considered reporting year, grade coverage, assessment, denominator, and indicator definition when determining whether comparisons were interpretable. Differences in populations or reporting scope were identified when relevant.

Contextual comparisons therefore answer a descriptive question: *How does the observed result sit relative to a locally or substantively relevant external benchmark?* They do not provide a counterfactual estimate of how the same learners would have performed in another high school model and should not be interpreted as evidence that CBL caused differences between groups.

Examining Relationships Between Competency and Academic Measures

Where compatible learner-level measures could be linked, Spearman rank correlations were used to examine relationships between local competency evidence and conventional academic indicators. Spearman correlation evaluates whether higher values on one measure tend to correspond with higher or lower values on another without requiring a linear relationship between the original scores. Nonparametric statistics were selected due to the overall positive skew observed across outcomes, as well as to be more appropriate for smaller sample sizes.

Each analysis was restricted to learners with usable records for both measures. The relevant competency definition, academic measure, reporting period, and paired-record sample size therefore determine the population to which each correlation applies.

Positive correlations indicate that learners with higher values on one measure tended to have higher values on the other. They do not establish that competency attainment caused academic performance, that academic performance caused competency attainment, or that the two measures assess identical constructs.

Observed associations may also be affected by restricted score variation, selective assessment participation, missing records, or incomplete linkage. Correlation results are therefore reported with the relevant paired-record sample sizes and interpreted in relation to the specific measures involved.

Exploratory Group Comparisons

Where the available data and subgroup sizes supported interpretation, exploratory analyses examined whether outcome distributions differed across demographic groups. Nonparametric procedures were used where measures were ordinal, distributions were skewed, or sample sizes made parametric assumptions less appropriate.

Kruskal–Wallis tests were used to examine differences in outcome distributions across multiple groups. Where follow-up comparisons were warranted, Dunn pairwise procedures were used to examine which groups differed.

These analyses were exploratory and remained secondary to the study’s descriptive outcome reporting. Statistical differences do not explain why groups differed, and nonsignificant findings do not establish equivalence between groups. Interpretation also depends on subgroup size, sample composition, missingness, and the meaning of the outcome being compared.

The site profiles therefore emphasize descriptive counts, percentages, and distributions and do not necessarily reproduce every exploratory statistical comparison conducted during analysis.

Interpretation and Analytic Limitations

Several features of the data bound the interpretations supported by the outcome analyses.

First, local competency measures differed across participating high schools in content, scoring, coverage, and purpose. These differences support meaningful within-high-school analyses but do not support direct ranking of high schools using a common competency outcome.

Second, outcome coverage varied within high schools. Some measures applied only to particular grades, tested learners, graduates, course participants, or learners with documented experiences.

Analyses combining competency and academic measures necessarily represent the subset of learners with usable information in both sources. Missing records, unmatched identifiers, small subgroups, and restricted score variation further affect the precision and scope of those analyses.

Third, most outcome analyses describe the 2024–25 school year, while the learner-experience surveys, classroom observations, and focus groups were conducted primarily in spring 2026. The outcome data therefore provide retrospective descriptive context for participating high schools rather than outcomes measured concurrently with the study’s learner-experience measures.

Fourth, contextual reference groups were not matched to participating high schools, and correlations do not isolate causal processes. The analyses do not account for all differences among learners, high schools, or external comparison populations and provide limited evidence about longer-term outcomes such as postsecondary persistence or completion.

Final findings are based on the verified analytic records, measure definitions, populations, and calculations used for each reported result rather than provisional availability inventories or undifferentiated master-file counts.

Within these boundaries, the analyses support descriptive interpretation of learner progress and performance within participating high schools and, where compatible learner-level data were available, examination of associations between local competency evidence and conventional academic measures. They do not support causal attribution, direct cross-site ranking of competency outcomes, or conclusions that the observed academic outcomes were produced by learners’ measured CBL experiences.

Appendix I.

Student Survey Instrument

The student survey was administered electronically through Qualtrics. The instrument below includes the respondent-facing questions, items, and response scales. Formatting has been standardized for publication. Construct labels are included to show how items were organized for analysis.

Section 1 of 7: Learning Expectations in Your School

In this section, we're going to ask about the competencies, concepts, and skills you are expected to learn at your school. Try to answer in general, not just for one teacher or class.

Clear Articulation

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I use my school's competencies to plan what I'll work on in different classes.
- I use my school's competencies to plan learning experiences outside my classes.
- I recognize when I'm demonstrating my school's competencies in different classes.
- I recognize when I'm demonstrating my school's competencies in non-school-related parts of my life.

Developmental Progression

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I choose learning experiences that help me reach higher levels of my school's competencies over time.
- When I move to a higher level of a competency, I can explain what I did differently to improve.

How true is the following of you?

Response scale: 1: Not at all true, 3: Somewhat true, 5: Very true

- I can explain what competency level I'm working toward.
- I can explain how close I am to mastering the competency level I'm working toward.

Connection to Learning Targets

Response scale: 1: Not at all true, 3: Somewhat true, 5: Very true

- I can explain what concepts and skills I'm working on in each of my classes.
- I can explain how close I am to mastering the concepts and skills I'm working on in each of my classes.
- I can explain the requirements for each level of a concept or skill.

Universal Application of Expectations

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I use a blueprint or rubric to understand how the concepts or skills I'm working on connect to a broader competency.
- When working on a competency or skill, I use a rubric that describes the mastery criteria.

How much do you agree or disagree with the following statements?

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- I expect teachers and other adults teaching a competency or skill to use the same criteria for evaluating my work.
- People teaching a competency or skill use the same criteria for evaluating my work across my classes and other learning experiences.

Clear Articulation

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- The competencies I'm working on will help prepare me for my future.

Section 2 of 7: Assessment in Your School

In this section, we're going to ask about how your knowledge and skills are assessed at your school. Try to answer in general, not just for one teacher or class.

Demonstration Location

How much do you agree or disagree with the following statements?

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- Only the people at my school can determine if my knowledge and skills meet my school's mastery criteria.
- Assessments from outside my school count as evidence that I have met my school's mastery criteria.
- When I demonstrate a competency or skill outside of school, people outside my school, such as supervisors, mentors, and coaches, can determine if it meets my school's mastery criteria.

Clear Calibration

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- No matter who assesses my work, the results consistently show whether I have met my school's competency criteria.
- No matter how my work is assessed, the results consistently show whether I have met my school's competency criteria.

Rigorous Application of Learning

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I demonstrate mastery of competencies by applying what I've learned to real problems or situations in my community.
- My assessments require me to demonstrate mastery through performance tasks that are meaningful.
- My assessments require me to combine several concepts or skills to address real problems or situations.

Section 3 of 7: Understanding Progression in Your School

In this section, we're going to ask about how you move forward in your learning at your school. Try to answer in general, not just for one teacher or class.

Mastery-Based Progression

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I keep working on a concept or skill until I demonstrate mastery.
- When I demonstrate mastery of a concept or skill, I move on to the next skill as soon as I'm ready.
- I keep working on a competency or skill until my work meets the mastery criteria, even if it takes extra time.
- If my work does not yet meet the mastery criteria, I spend more time improving it until it does.

Iterative Purpose

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- Before submitting my work for final credit, I revise it or practice more to show a deeper understanding.
- I retake assessments when necessary to demonstrate competency.
- I demonstrate my knowledge and skills through multiple drafts or assessment attempts.

Clear Calibration

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I use assessment results to decide whether to learn more about a concept, improve at a skill, or move on to the next competency.

Section 4 of 7: Understanding Growth in Your School

In this section, we're going to ask about how you keep track of and share your growth with others. Try to answer in general, not just for one teacher or class.

Transparency

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I keep a tracker, portfolio, or other record that organizes examples of my work by the competencies or skill levels they demonstrate.
- I check my progress tracker or portfolio to see how close I am to meeting the criteria for a competency or skill.
- I share my progress tracker or portfolio with teachers, supervisors/mentors, or classmates so they can review my work and give feedback on my competency or skill level.

Section 5 of 7: Understanding What's Important to You

In this section we're going to ask about what's important to you when you're learning. Try to answer in general, not just for one teacher or class.

Mastery Goal Orientation

How true is the following of you?

Response scale: 1: Not at all true, 3: Somewhat true, 5: Very true

- It's important to me that I learn a lot of new concepts this year.
- One of my goals in class is to learn as much as I can.
- One of my goals is to master a lot of new skills this year.
- It's important to me that I thoroughly understand my classwork.
- It's important to me that I improve my skills this year.

Performance Goal Orientation

Response scale: 1: Not at all true, 3: Somewhat true, 5: Very true

- It's important to me that other students in my class think I am good at my classwork.
- One of my goals is to show others that I'm good at my classwork.
- One of my goals is to show others that classwork is easy for me.
- One of my goals is to look smart in comparison to the other students in my class.

- It's important to me that I look smart compared to others in my class.

Section 6 of 7: Applied Learning

In this section, we're going to ask about learning activities you work on in your classes and outside of school. Try to answer in general, not just for one teacher or class.

Personalization

How often do you experience the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- My teachers consider my interests when deciding what I will work on.
- My teachers and I decide together what I will work on.
- If the work is too easy, my teachers will find a way to give me a challenge.
- Most of the activities in my classes are challenging, but not too challenging to complete.

Project-based Learning

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- Most of my time in my classes is spent working on a project.
- I can take ideas I've learned in one class and use them on my projects in other classes.
- My grade in my classes is determined mostly by the projects I work on.
- The projects I work on in one class also count towards my grades in other classes.
- I learn in my classes by completing projects.
- My teachers set the guidelines for the projects I work on in their classes.
- I know exactly what I am trying to learn when I work on projects in my classes.

Section 7 of 7: A little bit about you

You're almost done! We'd like to learn a bit more about you so we can help make sure that all students have a good experience at your school.

1. What school do you go to?

[Dropdown of all schools in their district]

2. What is your name?

3. How old are you?

- a. Younger than 11
- b. 11
- c. 12
- d. 13
- e. 14
- f. 15
- g. 16
- h. 17
- i. 18
- j. 19
- k. Older than 19

4. What grade are you in?

- a. My school doesn't use grade levels
- b. 6th grade
- c. 7th grade
- d. 8th grade
- e. 9th grade
- f. 10th grade
- g. 11th grade
- h. 12th grade
- i. Other

5. How long have you been going to this school?

- a. Less than 1 year
- b. 1-2 years
- c. 3-4 years
- d. More than 4 years

6. What is your gender?

- a. Female
- b. Male
- c. Other gender
- d. I don't want to share this information

7. What is your race? (Select all that apply.)

- a. American Indian or Alaskan Native
- b. Asian
- c. Black or African American

- d. Hispanic or Latino
- e. Middle Eastern or North African
- f. White or Caucasian
- g. Multiracial or Biracial
- h. A race not listed here
- i. I don't want to share this information

8. Do you speak a language other than English at home?

- a. Yes
- b. No
- c. I don't want to share this information

9. Do you receive any extra support to help you in school, like an IEP or 504 plan?

- a. Yes
- b. No
- c. I'm not sure
- d. I don't want to share this information

10. What grades did you get last year?

- a. Mostly D's or F's
- b. Mostly C's
- c. Mostly B's
- d. Mostly A's
- e. I'm not sure
- f. I don't want to share this information
- g. My school doesn't use grades

Appendix J.

Teacher Survey Instrument

The teacher survey was administered electronically through Qualtrics. The instrument below reproduces the survey questions, items, definitions, and response scales verbatim. Formatting has been standardized for publication. Construct labels are included to show how items were organized for analysis.

Key Terms Used in the Survey

- **Competencies** are transferable knowledge, skills, and dispositions that students are expected to demonstrate.
- **Mastery and Competency** are used interchangeably for when students demonstrate that they met competencies or learning outcomes at the pre-identified level, although the way they demonstrate mastery can be vastly different.
- **Courses** (also known as a “class”, “seminar”, or “learning experience”) are opportunities for students to participate in a set of learning tasks and other activities designed to advance their knowledge and skills in a specific subject or skill area.
- **Teachers** are adults who facilitate student learning in school settings. This includes adults who may be called “educators” or “advisors.”
- **Credits** are units of measure that indicate whether a student has successfully satisfied course requirements and/or demonstrated mastery of a set of learning outcomes.

About You and Your School

1.1. What is your school?

[Dropdown list of schools]

1.2. Are you a teacher or staff member who has direct responsibility for instructing, assessing, grading, and facilitating learning of students?

- Yes
- No

[Skip Logic] If Q1.1 = “No,” skip to end-of-survey message

1.3. What is/are your primary area(s) of teaching / facilitating learning? Select all that apply.

- Elementary education
- English language arts

- Career/technical education
- Foreign language
- Instructional coach
- Mathematics
- Physical education
- Science
- Social studies
- Special education
- The arts (e.g., music, visual art)
- Other(_____)

1.4. What grades of students do you teach? If your school doesn't assign grade levels, please indicate the ages of the students you teach, shown in parentheses. Select all that apply.

- Kindergarten (age 5 or younger)
- 1st grade (or age 6)
- 2nd grade (or age 7)
- 3rd grade (or age 8)
- 4th grade (or age 9)
- 5th grade (or age 10)
- 6th grade (or age 11)
- 7th grade (or age 12)
- 8th grade (or age 13)
- 9th grade (or age 14)
- 10th grade (or age 15)
- 11th grade (or age 16)
- 12th grade (or age 17 or older)

1.5. How many years have you been teaching at this school?

- 3 years or less
- 4-9 years
- 10-14 years
- 15 or more years

1.6. How many years have you been teaching at this school and any previous schools combined?

- 3 years or less
- 4-9 years
- 10-14 years
- 15 or more year

Section 1 of 4: Learning Expectations in Your School

In this section, we're going to ask about the competencies, concepts, and skills students in your class are expected to learn. Try to answer in general, not just for one student, class, or advisory period.

Clear Articulation

How often do you observe the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- Students in my class use our school's competencies to plan their work across their classes.
- Students in my class use our school's competencies to plan their work in learning experiences outside their classes.
- My students can identify when the same competencies are being developed in different classes.
- My students can identify when the competencies from my school are being developed in other parts of their lives.

Developmental Progression

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- I choose learning experiences that help my students reach higher levels of my school's competencies over time.
- When my students move to a higher level of a competency, I can explain what they did differently to improve.

How true is the following of students in your class?

Response scale: 1: Not at all true, 3: Somewhat true, 5: Very true

- Students in my class can explain what competency level they're working toward.
- Students in my class can explain how close they are to mastering the competency level they're working toward.

Connection to Learning Targets

Response scale: 1: Not at all true, 3: Somewhat true, 5: Very true

- My students can explain what concepts and skills they're working on in my class.
- My students can explain how close they are to mastering the concepts and skills they're working on in each of their classes.

- Students in my class can explain the requirements for each level of a concept or skill.

How often do you observe the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- My students use a blueprint or rubric to understand how the concepts or skills they're working on connect to a broader competency.

Universal Application of Expectations

How often do you observe the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- When working on a competency, my students use a rubric that describes the mastery criteria.
- The same mastery criteria are used across my students' classes and learning experiences.

How much do you agree or disagree with the following statements?

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- Students in my class expect teachers and others teaching a competency or skill to use the same mastery criteria.

Clear Articulation

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- The competencies my students are working on help prepare them for their future.

Section 2 of 4: Assessment in Your School

In this section, we're going to ask about how your students' knowledge and skills are assessed at your school. Try to answer in general, not just for one student or class or advisory period.

Demonstration Location

How much do you agree or disagree with the following statements?

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- Only the people at my school can assess whether my students' knowledge and skills meet the school's mastery criteria.
- Assessments from outside my school count as evidence that my students have met my school's mastery criteria.
- People outside my school evaluate the work my students do outside of school to determine whether it meets the school's mastery criteria.

Clear Calibration

Response scale: 1: Strongly disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5: Strongly Agree

- No matter who assesses their work, the results consistently show whether students in my class have met my school's competency criteria.
- No matter how their work is assessed, the results consistently show whether students in my class have met my school's competency criteria.

Rigorous Application of Learning

How often do you observe the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- Students in my class demonstrate mastery of competencies by applying what they've learned to real problems or situations in their community.
- Students in my class demonstrate mastery through performance tasks.
- Students in my class demonstrate mastery by applying several concepts or skills to real problems or situations.

Section 3 of 4: Understanding Progression in Your School

In this section, we're going to ask about how students in your school move forward in their learning. Try to answer in general, not just for one student or class or advisory period.

Mastery-Based Progression

How often do you observe the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- My students keep working on a concept or skill until they've demonstrated mastery.

- When my students demonstrate mastery of a concept or skill, they move on to the next skill as soon as they're ready.
- My students keep working on a competency or skill until their work meets the mastery criteria, even if it takes extra time.
- If their work does not yet meet the mastery criteria, my students spend more time improving it until it does.

Iterative Purpose

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- Before submitting their work for final credit, students in my class revise it or practice more to show a deeper understanding.
- Students in my class retake assessments when necessary to demonstrate competency.
- Students in my class demonstrate their knowledge and skills through multiple drafts or assessment attempts.

Clear Calibration

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- Students in my class use assessment results to decide whether to learn more about a concept, improve at a skill, or move on to the next competency.

Section 4 of 4: Understanding Growth in Your School

In this section, we're going to ask about how students in your school keep track of and share their growth with others. Try to answer in general, not just for one student or class or advisory period.

Transparency

How often do you observe the following?

Response scale: 1: Never, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

- My students keep a tracker, portfolio, or other record that organizes examples of their work by the competencies or skill levels they demonstrate.
- My students check their progress tracker or portfolio to see how close they are to meeting the criteria for a competency or skill.
- My students share their progress tracker or portfolio with teachers, mentors, or classmates so they can review their work and give feedback on their competency or skill level.