

Connecting ECE Quality Measures to Later Outcomes

**Evaluating Alignment between Arkansas's
Early Care and Education Quality Measures
and Later Child Outcomes**

Hannah Denker, Ph.D.

University of Arkansas

April 1, 2026

Abstract: This report evaluates the extent to which early childhood quality measures used in Arkansas align with later student outcomes, including kindergarten readiness and early-grade academic performance. Drawing on developmental theory, empirical research, and a review of commonly used tools, the analysis finds that no single measure captures all dimensions of readiness. Observational measures, such as those focused on adult-child interactions, align with developmental processes that support long-term learning, while direct academic assessments more strongly predict later test performance but reflect narrower domains. Structural and environmental indicators contribute indirectly by shaping classroom conditions. The findings suggest that Arkansas's current measurement system includes many of the necessary components for alignment, but clearer guidance could further support consistent and appropriate use of these measures. The report identifies opportunities to clarify the purpose of each tool, strengthen alignment across birth-to-grade-2 systems, and support the use of complementary measures to inform policy and practice.

Keywords: Early care and education, Quality Rating and Improvement System (QRIS), Kindergarten readiness, Policy alignment

Table of Contents

I.	Executive Summary	4
II.	Purpose, Scope, and Guiding Questions	6
	Purpose.....	6
	Scope	6
	Guiding Questions.....	6
III.	Kindergarten Readiness as an Outcome	7
IV.	Defining Alignment as a Goal	12
V.	Evidence Review: What Types of Measures Align with Later Outcomes?.....	14
	Measures of Adult–Child Interactions	14
	Strength of Evidence and Appropriate Use	18
	Child Skill Screeners and Kindergarten Entry Assessments	20
	Comprehensive KEAs.....	21
	Developmental Checklists	22
	Direct Skill Screeners	22
	Strength of Evidence and Appropriate Use	24
	Observational Child Development Measures	26
	Teaching Strategies GOLD.....	27
	The Work Sampling System (WSS)	29
	The Ounce Scale	31
	Strength of Evidence and Appropriate Use	32
	Global Environmental and Structural Quality Indicators.....	34
	Teacher Qualifications and Experience	35
	Class Size and Child-Teacher Ratios.....	37
	Curriculum Implementation and Fidelity.....	38
	Global Environment & Administration (ERS & PAS).....	39
	Strength of Evidence and Appropriate Use	41
	Early-Grade Academic and Developmental Outcomes as Alignment Anchors	43

Academic Outcomes	43
Social-Emotional and Self-Regulation Outcomes	44
Health and Developmental Indicators.....	45
Strength of Evidence and Appropriate Use	46
VI. Arkansas’s Measurement Landscape.....	48
Policy Context in Arkansas.....	48
Mapping Arkansas Measures to Readiness Domains	51
VII. Alignment and Gaps in Arkansas’s Current Approach.....	55
Areas of Alignment.....	55
Use of process quality measures aligned with developmental theory	55
Integration of early-grade academic assessments.....	55
Multi-tool measurement system across early childhood and K–2	55
Areas of Potential Misalignment or Gaps	56
Unclear Guidance on the Intended Purpose of Different Measures	56
Use of CLASS® as a Proxy for Child Readiness or High-stakes Accountability. 56	
Variation in Implementation and Fidelity across Observational and Checklist- Based Tools	56
Narrow Domain Coverage of Academic Assessments	56
Development of a Formal Alignment Framework across Birth-to-Grade-2 Systems.....	57
Implications for QRIS Policy and Implementation.....	57
VIII. Recommendations Focused on Measurement and Alignment.....	61
IX. End Notes.....	64

I. Executive Summary

The purpose of this report is to evaluate the extent to which early childhood quality measures and related tools currently used in Arkansas align with later student outcomes, including kindergarten readiness and early-grade academic and developmental indicators. This report draws on developmental theory, empirical research, and a review of commonly used evaluation tools in Arkansas. From these sources, the report examines how well these evaluation measures capture aspects of early learning environments and child development that are linked to later outcomes.

Three conclusions guide this report. First, kindergarten readiness is a multidimensional and context-dependent construct, and no single measure captures all aspects relevant to later student success. Research indicates that later student success depends not only on early literacy and mathematics skills, but also on social-emotional development, self-regulation, approaches to learning, and physical well-being. Each of these domains requires different early care and education (ECE) evaluation tools to track.

Second, different types of measures align with later student outcomes in various ways. Direct academic screeners tend to show stronger associations with later test scores because they assess the same types of skills emphasized in early elementary accountability systems, whereas observational measures, like those focused on adult-child interactions, align more closely with the developmental mechanisms that support long-term learning. Structural quality indicators, like staff-child ratios, remain important, but their influence on child outcomes is generally indirect and operates through the classroom experiences they make possible.

Third, Arkansas currently has many of the necessary components of a birth-to-grade-2 aligned measurement system, and continued development of a unified framework clarifying how these tools relate to one another would further strengthen how they inform improvement and policy decisions across settings.

Based on this evidence, this report recommends clarifying the intended purpose of each measure used in the state’s Quality Rating and Improvement System (QRIS), strengthen cross-sector alignment across birth-to-grade-2, and use multiple complementary measures to evaluate quality, readiness, or effectiveness. Building on the findings of this alignment analysis, future work will build on this analysis to provide recommendations for ECE quality measures for inclusion in Arkansas’s QRIS.

II. Purpose, Scope, and Guiding Questions

Purpose

This report evaluates the alignment between Arkansas’s early care and education (ECE) quality measures and later student outcomes, specifically kindergarten readiness and early-grade academic performance. The primary objective is to determine whether the state’s current ECE quality definitions, and the tools used to measure them, reflect the specific developmental indicators that research identifies as key drivers of future success.

Scope

This analysis assesses system alignment rather than program efficacy. It is not a causal evaluation of specific interventions, nor is it a comprehensive audit of statewide student readiness. Instead, this report focuses on the specific measurement tools used within Arkansas's ECE and early elementary systems to determine:

- **Do the tools predict later success?** Do the ECE quality measures currently used in Arkansas accurately identify the skills and environments associated with positive outcomes later in school?
- **What is the correct use of these tools?** To what extent are they being applied in ways consistent with their original design?

Guiding Questions

To address these objectives, the report answers the following questions:

1. How does research define “kindergarten readiness,” and what does this imply for valid measurement?

2. Which ECE quality measures and tools show strong evidence of alignment with later academic and developmental outcomes?
3. To what extent do Arkansas's current assessment tools map onto these evidence-backed standards?
4. Where are the gaps or misalignments between what Arkansas's QRIS currently measures and the factors that are most important for driving student outcomes?

III. Kindergarten Readiness as an Outcome

Before evaluating specific measurement tools in later sections of this report, it is important to establish a shared understanding of what "kindergarten readiness" means. We focus on kindergarten readiness because it represents a widely recognized, policy-anchored milestone with validated measurement tools.¹

However, the usefulness of any measure of kindergarten readiness depends on how clearly the concept itself is defined. If the definition is too narrow, the resulting data will be incomplete; if the definition is ignored in practice, the data will be misleading. Accordingly, the following analysis uses a multidimensional framework to ensure that the assessment tools under review are evaluated against a holistic definition that includes multiple key dimensions in child development.

While definitions vary slightly from state to state, we use a multidimensional framework originally established by the National Education Goals Panel (NEGP) and reaffirmed by a 2008 report from the National Academies of Sciences, Engineering, and Medicine.² This consensus

holds that readiness cannot be captured by a single test score or academic metric, not only because of limitations in available tools, but also because readiness is itself a multidimensional construct spanning distinct developmental domains that are important to have separate information about simultaneously.³ Instead of readiness as a single test score, it has been described as a construct comprising five essential domains:⁴

- **Physical Well-Being and Motor Development:** Health status, growth, and physical abilities.
- **Social and Emotional Development:** The ability to interact with others and manage emotions.
- **Approaches Toward Learning:** Curiosity, persistence, and the predisposition to engage in learning.
- **Language Development:** Verbal skills, emergent literacy, and communication.
- **Cognition and General Knowledge:** Basic mathematical and literacy concepts, in addition to awareness of the surrounding world.

Arkansas aligns with the vast majority of U.S. states in how it officially defines readiness. According to a 2019 analysis by the Educational Testing Service (ETS), most states with formal definitions, including Arkansas, have adopted an *ecological* perspective.⁵ This means Arkansas policy officially views readiness as an interaction between the child and their environment (families, schools, and communities), rather than a *nativist* view that treats readiness solely as a fixed trait of the child.

Arkansas’s definition explicitly states: *“School-ready children have the social and academic knowledge, skills, and behaviors for school success and lifelong learning. School readiness occurs when families, schools, and communities support and serve ALL children.”*⁶

Even with a robust definition, actually measuring readiness is difficult. Research consistently shows that early childhood screenings provide only a snapshot of a rapidly developing child.⁷ Because children grow and learn in bursts during these early years, a test score at age four does not always predict how well a student will perform by age seven or eight. In fact, research indicates that these early scores are most effective when used as indicators of current support needs rather than as long-term predictions of academic potential.⁸ This unevenness means that while readiness data is vital for identifying current needs, it should be used with caution when trying to predict a child's longer-term potential.

Policymakers must also navigate a tension between the developmental priorities of educators and the expectations of families. While teachers overwhelmingly identify learning-related skills, such as self-regulation and attention, as the biggest growth opportunities for incoming students,⁹ parents hold more varied views. Research on low-income Latino families found that parents in that population often emphasized discrete academic skills, defined here as specific, easily observed early literacy and numeracy tasks such as counting and letter recognition, as primary indicators of readiness.¹⁰ However, a broader review of 71 studies across 17 countries found that parents frequently prioritize both interpersonal skills and academic development, with cultural context shaping the relative emphasis placed on each.¹¹ These expectations exist within a broader shift in which kindergarten classrooms have become

substantially more academically oriented over the past two decades, with increased time devoted to literacy and math instruction and greater use of standardized assessments.¹² Bridging the gap between increasing demand for early academic skills and the developmental competencies that support sustained learning is a key policy opportunity. Early childhood and elementary education systems can meet the demand for academic readiness by ensuring access to high-quality curricula and instruction that promote literacy and numeracy while also prioritizing the developmental skills that make sustained learning possible.

When measurement tools face pressure to focus too narrowly on discrete skills, they risk measuring opportunity rather than potential. Research indicates that significant disparities in access to early learning resources, such as high-quality child care and literacy materials in the home, exist between socioeconomic groups long before kindergarten begins.¹³ If a measurement system ignores these environmental factors, it risks penalizing children from lower-income and minority backgrounds. In these cases, a "not ready" label often reflects a lack of prior access to high-quality early learning experiences rather than a lack of ability to learn.¹⁴

Just as child-level readiness scores can misrepresent a child's potential, they are also problematic metrics for evaluating the efficacy of individual ECE programs. Because readiness is an accumulation of experiences across home, community, and multiple care settings, it is methodologically difficult to attribute a child's kindergarten entry score solely to their most recent ECE provider.¹⁵ Using point-in-time readiness scores for high-stakes accountability, such as in funding decisions, risks punishing providers who serve the highest-need populations while failing to account for differences in opportunity to learn after kindergarten entry and the well-

documented, but not yet understood, *fade-out effects* often seen when early gains are not sustained in the K-12 system.¹⁶

Given the complexities of the multidimensional nature of development, the difficulty of prediction, and the risks of narrow measurement, it is insufficient to simply ask if a tool is *valid* (i.e., it measures what it is intended to measure). Research guidance on this point is consistent; readiness assessments are most appropriately used to identify children who may benefit from additional instructional support at school entry and to inform classroom planning, rather than to rank programs, generate public report cards on ECE programs, or serve as the basis for high-stakes funding decisions.¹⁷ The following section defines what "alignment" means in the context of Arkansas's system and how we can ensure our metrics match the comprehensive view of readiness that reflects children's experiences across home, early care, and school settings.

A note on terminology: several related terms appear throughout this report and warrant upfront clarification. Cognitive skills refer broadly to mental processes involved in learning, reasoning, and problem-solving. Executive functions are specific neurocognitive capacities, including working memory, inhibitory control, and cognitive flexibility, that regulate attention, emotion, and behavior. Self-regulation refers to a child's ability to manage their emotions, attention, and actions, and is best understood as a downstream product of developing executive function rather than a standalone behavioral skill that can be trained in isolation. Social-emotional development encompasses the broader capacity to understand and manage emotions and build relationships with others. This report uses the umbrella term "behavioral skills" to

mean observable expressions of these underlying capacities, not behaviors targeted through compliance-based training.

IV. Defining Alignment as a Goal

Because kindergarten readiness is defined in multiple ways across research, policy, and practice, this report evaluates alignment through two distinct lenses:

- **Empirical Alignment:** The degree to which a quality measure statistically predicts or is related to a child’s later learning outcomes (e.g., do high ratings on Tool X predict high scores on Test Y?).
- **Conceptual Alignment:** The degree to which a measure captures the developmental mechanisms, like responsive interactions or self-regulation, that theory identifies as drivers of learning, even if those mechanisms are difficult to capture in standardized testing.

To evaluate this alignment, the report analyzes the specific categories of measurement tools that have been recently used or are currently available in Arkansas:

- **Measures of Adult–Child Interactions:** Instruments that assess the quality of teacher–student exchanges (e.g., Classroom Assessment Scoring System® [CLASS®]).
- **Child Assessments:** Tools that evaluate student progress, ranging from direct skill screeners (e.g., ATLAS) to observational developmental instruments (e.g., Teaching Strategies GOLD [TS GOLD]).

- **Global & Structural Indicators:** Metrics related to the physical environment, administration, and program characteristics (e.g., Environment Rating Scales [ERS], Child-Staff ratios).

It is critical to distinguish between alignment and accountability. Finding that a measure is aligned with later outcomes does not automatically make it suitable for high-stakes decision-making. Many early childhood measures were designed for formative purposes, such as guiding professional development, rather than for holding programs accountable or evaluating staff performance.¹⁸

Consequently, a measurement tool may be highly useful for improving classroom practice even if its statistical relationship with later standardized test scores is small to medium in magnitude. This can be particularly true when the measurement tool captures short-term, specific instructional processes while test scores reflect more long-term, complex outcomes. Alignment, therefore, also requires attention to whether each tool is being applied for purposes consistent with its design and evidence base. For example, it is important to consider whether an instrument developed for coaching and professional development is being used in that capacity or extended to serve as a program rating or accountability metric. The central policy question is not just whether a measure predicts a child's test score, but whether it captures specific skills the state values,¹⁹ both *constrained skills*, which are discrete and finite early academic tasks that can be fully mastered and have a defined ceiling (e.g., letter naming), and *unconstrained skills*, which develop gradually over time and support learning across contexts (e.g., vocabulary and self-regulation).

Therefore, in this report, alignment is treated as a descriptive metric that can guide system improvement, not as a gatekeeping threshold. Our analysis focuses on how existing measures can be used to support child development and inform improvement, rather than proposing broad changes to the current measurement infrastructure.

V. Evidence Review: What Types of Measures Align with Later Outcomes?

This section summarizes how different early childhood measures relate to later outcomes, including kindergarten readiness, early academic achievement, and broader developmental indicators. Evidence includes developmental theory, longitudinal research studies, and statistical meta-analyses (i.e., studies that summarize statistical findings across many independent research evaluations) of early childhood measurement tools. For each category, the section summarizes theoretical alignment, empirical evidence, appropriate policy uses, and limitations.

Research consistently shows that improving child outcomes is genuinely difficult, and even a result characterized as "small" can represent a meaningful gain when it is replicated across many children and classrooms. Throughout this report, small, moderate, large, and null (i.e., no relationship) describe the *size* of a finding; weak/strong, inconsistent/consistent, and null (i.e., no relationship) describe our *confidence* in it. A small but consistent finding is still meaningful; a large but inconsistent one warrants caution.

Measures of Adult–Child Interactions

Measures of adult–child interactions, including the Classroom Assessment Scoring System® (CLASS®), are grounded in developmental and sociocultural theory emphasizing the role of responsive, instructionally rich interactions in supporting children’s learning and

development.²⁰ Building on this framework, CLASS[®] captures dimensions of emotional support, classroom organization, and instructional support, which are theorized to influence language development, self-regulation, and early academic skills.²¹ Importantly, CLASS[®] is best understood as a useful starting point for understanding the quality of adult–child interactions in classrooms, rather than a comprehensive measure of the full range of instructional quality or student learning.

Empirical evidence linking CLASS[®] domains to child outcomes suggests some statistically significant positive associations, although they are often small and inconsistent, particularly at the individual child level. Recent meta-analyses, or studies that summarize statistical findings across many independent research evaluations, report small to moderate positive relationships, on average, between CLASS[®] scores and children’s language and literacy outcomes. These meta-analyses find smaller, and often weaker, relationships for mathematics, executive function, social development, attention, and behavior.²² Other meta-analytic evidence confirms that associations between CLASS[®] instructional support and children’s language outcomes are statistically significant, though small in size.²³ Positive relationships tend to be more consistent for the instructional support domain and language-related outcomes than for other domains, though the relationship tends to be small, on average.²⁴

Large studies that take into account differences in child and classroom characteristics, along with replication studies to see if findings consistently hold true in different contexts or more recent times, frequently find no or small statistical associations between CLASS[®] scores and gains in children’s academic, language, and executive function skills.²⁵ This is particularly

true in settings where classroom quality is relatively high overall and variation in the scores is limited.²⁶ These findings suggest that in settings with limited variation in CLASS[®] scores, the scores alone may not strongly predict individual children's readiness or short-term learning gains.

However, two studies point to larger, positive results. First, a study of Louisiana's CLASS-based QRIS found that continuous CLASS[®] scores significantly predicted children's composite learning gains across math, language, and literacy measures, and that children in classrooms rated in the middle and top CLASS[®] categories showed meaningfully larger gains than those in the lowest-rated classrooms.²⁷ Notably, these results were obtained using CLASS[®] scores collected by local observers under routine QRIS conditions rather than by a research team, as is often the case in empirical evaluations. The same study also found that a CLASS-only rating system performed comparably to a simulated multi-indicator system in distinguishing program quality, supporting the case for streamlined approaches to quality measurement, although this study is unique in this conclusion.

Another study found CLASS[®], particularly its domains measuring teacher–child interactions, is linked to later student outcomes, but the strength and direction of these relationships vary by domain and subject.²⁸ Overall, higher classroom quality as measured by CLASS[®] is related to improvements in academic outcomes, especially in reading, with effects that can persist beyond pre-kindergarten. Notably, the classroom organization (CO) domain shows the most consistent positive relationship with later achievement: higher CO scores in pre-kindergarten predict better reading performance through kindergarten, first, and second grade, suggesting that well-managed, structured classrooms support sustained academic growth. In

contrast, the other CLASS® domains, emotional support (ES) and instructional support (IS), show weaker or inconsistent statistical relationships, and in some cases even negative relationships with outcomes. Together, these results indicate that not all aspects of observed classroom quality translate equally into academic gains.²⁹

The difference in magnitude between these findings and the smaller and weaker relationships observed in other studies likely reflects differences in the children in the samples, CLASS® score variability (or lack thereof), analytic approach, and study context, emphasizing that how predictive CLASS® is for child outcomes may depend substantially on how and where it is implemented. Importantly, the limited range of CLASS® scores observed in some studies that might limit the predictive power of CLASS scores may itself be a marker of system success. In other words, when programs consistently maintain high interaction quality, the tool has served its improvement purpose even though its predictiveness of student outcomes diminishes.

In sum, these results do not imply that adult–child interactions are unimportant for children’s development. Rather, they reflect the fact that CLASS® was designed as a measure of classroom process quality, not as an outcome assessment. Evidence indicates that specific interactional processes, such as responsive teaching and language-rich interactions, show small to moderate positive relationships with children’s cognitive and social domains.³⁰ *The tool is therefore most appropriately used to support professional learning, coaching, and continuous quality improvement by identifying instructional practices and interaction patterns that research has linked to positive developmental trajectories for children.*³¹

From an alignment perspective, the evidence indicates that CLASS® aligns more closely with developmental mechanisms and learning processes than with direct prediction of later child outcomes.³² This reflects both the design and scope of the measure: CLASS® captures the quality of adult–child interactions at the classroom level, but it does not directly assess the full range of instructional content, curriculum implementation, or individual child learning. This distinction has important implications for policy. When used for improvement and professional development, CLASS® is well aligned with the goal of strengthening early learning environments. When used as a proxy for child readiness or as a high-stakes accountability measure, alignment with later student outcomes is more limited.³³

Strength of Evidence and Appropriate Use

Alignment with later child outcomes.

Across studies, CLASS® scores tend to show small to moderate positive associations with children’s later outcomes, with the most consistent relationships observed for language and literacy skills. These associations vary by CLASS® domain, context, and study design.

Summary judgment.

Evidence on the association between CLASS® and longer-term child outcomes is mixed. Some studies, particularly those with greater variability in classroom quality, find positive moderate associations with learning gains, especially in language and literacy. Other well-controlled studies, particularly in settings where quality as measured by CLASS® is relatively uniform, find small or no relationships with child outcomes. Additionally, smaller, more

inconsistent relationships are observed for mathematics, executive function, and behavioral outcomes. The weight of evidence supports CLASS® as a meaningful indicator of classroom process quality rather than as a direct predictor of individual child outcomes.

Appropriate policy use based on evidence.

CLASS® is best aligned with improvement-oriented purposes, including professional learning, coaching, and identification of interactional practices that consistently show positive associations with early developmental outcomes, particularly in language and literacy. Ultimately, CLASS® seems to be most effective when used to guide professional growth and instructional improvement rather than as a standalone tool to evaluate individual child outcomes. It identifies pathways of growth for early childhood educators that can lead to quality, even when those paths do not immediately register on a standardized test. Evidence supports its use as a measure of classroom process quality, rather than as a direct measure of child readiness or a standalone program accountability metric.

Key limitations affecting alignment.

Inconsistent Effects: In meta-analyses, or studies that summarize effects across many independent studies, relationships between CLASS® scores and child outcomes are often small to moderate in size. Other well-controlled studies and replication efforts find no measurable statistical link between scores and short-term gains in child outcomes. This pattern may be partly explained by a level-of-analysis mismatch: CLASS® is a classroom-level measure that captures the average quality of interactions across a group of children. Using it to predict variation in

outcomes for individual children asks it to explain differences that are also shaped by each child's unique prior development, family context, and experiences outside the classroom, none of which CLASS® is designed to capture.

Limited Variation: In many publicly funded pre-k settings, CLASS® scores may cluster in a narrow range (limited variation), making it difficult to statistically detect their impact on child outcomes. For example, in programs that already maintain a high standard of adult-child interaction quality, scores often cluster at the top of the scale. This “ceiling effect” means the tool has already done its job of ensuring baseline quality, but it makes the scores less useful for predicting small differences in individual children’s development.

Specific Design: The tool was designed to measure classroom environments, not to serve as a direct proxy for individual child skills. CLASS® can help identify classrooms or programs that may not yet provide a minimum level of interaction quality, allowing targeted supports and resources to be directed to strengthen practice.

Child Skill Screeners and Kindergarten Entry Assessments

Because kindergarten entry assessments are often used as a benchmark for children’s readiness at school entry, they provide an important point of reference for evaluating how well early childhood quality measures align with later outcomes. To evaluate children’s status at school entry, states typically use one of three approaches: comprehensive Kindergarten Entry Assessments (KEAs), developmental checklists, or direct skill screeners. Kindergarten Entry Assessments and developmental checklists prioritize capturing how children function in everyday classroom contexts, including social, physical, and emotional behaviors alongside

academic skills, even if results in less standardized and less directly comparable data across settings than direct skill screeners.³⁴ In contrast, direct skill screeners differ from observation in that they prioritize statistical precision, focusing on effectively measuring discrete skills, that is, skills that can be objectively measured for an individual child.

Arkansas has used all three approaches at different points in its policy history, creating a range of alignment data in which different tools capture distinctive aspects of "readiness." This complexity is compounded by the fact that different funding streams, including federal Head Start requirements, Arkansas Better Chance (ABC) program expectations, and child care subsidy administration, have historically carried distinct reporting mandates.

Comprehensive KEAs

For many years, Arkansas implemented the Qualls Early Learning Inventory (QELI) as its formal Kindergarten Entry Assessment. Administered by kindergarten teachers at the start of the school year, the QELI was a teacher-completed observation tool that assessed students across six domains, including physical development and social adjustment.³⁵

- **Alignment Function:** Because the QELI relied on teacher observation of broad behaviors, it aligned conceptually with the "whole child" philosophy of early childhood development.
- **Limitation:** Like many observational KEAs, it was subject to teacher subjectivity and "reference bias," a pattern documented in the broader literature, where ratings varied significantly depending on the teacher's internal standards rather than the child's objective

ability (i.e., called “low reliability” in statistical parlance).³⁶ It was eventually phased out in favor of faster, more standardized metrics. To our knowledge, Arkansas-specific statistical data on the QELI's reliability are not available in the published literature.

Developmental Checklists

To retain a focus on the whole child, Arkansas continues to informally use the Kindergarten Readiness Indicator Checklist (KRIC). Unlike the ATLAS screener, which is administered in kindergarten, the KRIC is completed by families and early childhood educators during the ECE-to-kindergarten transition period, making it an ECE-side transition tool rather than a formal entry measure. It documents readiness across 38 indicators spanning domains that broadly correspond to the Arkansas Child Development and Early Learning Standards (CDELS), including interpersonal skills and physical health,³⁷ though a formal published alignment analysis between the KRIC and the current CDELS standards is not available.

- **Limitations:** Research indicates that checklist-based tools are also subject to "reference bias," where ratings reflect not only children's behaviors but also differences in how observers interpret and score those behaviors. Recent empirical analyses suggest that teachers may systematically rate dual language learners lower on observational readiness scales compared to their performance on direct assessments, raising equity concerns for using tools like the KRIC as standalone metrics for diverse populations.³⁸

Direct Skill Screeners

Currently, Arkansas policy mandates the use of a direct skill screener (i.e., the ATLAS K–2 assessment suite) to identify Kindergarten students needing academic intervention under the

LEARNS Act. This represents a distinct shift in the state's Kindergarten readiness measurement philosophy.

- **Focus on Constrained Skills:** Unlike the QELI or KRIC, modern screeners focus heavily on *constrained skills*, defined as discrete, measurable items like letter naming, phonemic awareness, and counting. Empirical meta-analyses, or studies that use a statistical method to summarize findings from multiple independent, similar studies, confirm that early academic skills are among the strongest available predictors of later achievement, generally outperforming broad social-emotional ratings in predicting later academic assessment scores.³⁹
- **Limitations:** While screeners provide reliable data for "instructional triage," they also reflect the well-documented 'fade-out' phenomenon observed in local evaluations. For instance, the NIEER evaluation of the Arkansas Better Chance (ABC) program found that while ABC produced significant initial gains in literacy and math (detected by screeners), these advantages narrowed significantly in later elementary grades.⁴⁰ This pattern may reflect that later schooling reduced gaps in learning opportunities, as well as differences between ECE and early elementary instructional approaches.⁴¹ In addition, because screeners focus on discrete, measurable skills, they are less well suited to capturing more complex, unconstrained skills (e.g., language development and higher-order thinking), which likely play an important role in sustained learning over time.

The existence of both KEAs and ATLAS creates a *measurement bifurcation*, meaning that Arkansas uses two different types of tools to evaluate readiness, each designed for a different

purpose. Observational KEAs, which are aligned with the CDELS, focus on the developmental aspects of the ECE system, such as children’s approaches to learning. In contrast, the ATLAS Kindergarten screener measures discrete academic skills more directly tied to later accountability measures. This means a child can show high levels of learning-related behaviors (e.g., persistence, following directions, etc.) while still requiring targeted academic support. When interpreted together, these measures provide a more complete picture of children’s readiness, showing that ECE programs may be building foundational learning behaviors even as the ATLAS screener identifies areas for targeted academic support.⁴²

Strength of Evidence and Appropriate Use

Alignment with later child outcomes.

Direct skill screeners, like ATLAS, provide the most consistent prediction of later state test scores because they measure the specific academic building blocks, like letter sounds and number recognition, that form the basis of later state standards. In contrast, observational KEAs (QELI/KRIC) are less effective at predicting future test scores but provide a more comprehensive view of a child’s developmental progress, which may support their long-term classroom success, even though they are not designed to predict individual academic outcomes.

Summary judgment.

Direct skill screeners provide the strongest available prediction of later academic outcomes because they measure the same constrained skills that form the basis of early elementary standards. However, this predictive strength is narrow as screeners do not capture the

interpersonal, self-regulatory, and physical development domains that research identifies as important for sustained developmental and school success. Observational KEAs and developmental checklists offer broader coverage but are limited by rater subjectivity and lack the psychometric precision needed for child-level accountability. The weight of evidence supports treating these tools as having different purposes, e.g., screeners for instructional triage and early academic intervention, and observational tools for transition planning, family engagement, and monitoring developmental progress across the domains that screeners cannot capture.

Appropriate policy use based on evidence.

These tools are best treated as complementary, with screeners guiding kindergarten instruction and observational KEAs supporting the ECE-to-kindergarten transition. However, research cautions against using point-in-time scores from either tool type to evaluate program efficacy, as doing so risks penalizing programs that serve high-need populations without accounting for the cumulative nature of children's developmental experiences across settings.⁴³

Key limitations affecting alignment.

Narrow Scope (Screeners): The ATLAS kindergarten screeners are designed to flag early academic risk (e.g., early reading difficulty). Based on our review of ATLAS content specifications, literacy items are largely based on end-of-pre-k expectations while the math items function more as a pre-test of early kindergarten standards. As a result, these screeners reflect differences in prior opportunity to learn as well as potential early academic difficulties. The academic focus misses interpersonal, self-management, and physical development skills that

ECE programs are designed to foster and that likely support later learning. Finally, because these tools focus on constrained skills (which have a ceiling once mastered), they are better for early triage but not designed for capturing unconstrained skills, like critical thinking, that are also important for long-term learning.

Reference Bias (Checklists): Variability in how different parents and teachers rate items on observational KEAs reduces their usefulness in statewide comparisons, particularly for dual language learners.

Data Silos: The lack of a unified data system connecting what KEAs measure in ECE to what the ATLAS screener measures in kindergarten prevents the state from examining how early developmental strengths are related to later student success. Equally important, this data silo prevents ECE programs from accessing early-grade outcome data for their former students. This could provide a direct signal for identifying where ECE practices are effectively preparing children and where there are opportunities to strengthen alignment with what kindergarten instruction will require.

Observational Child Development Measures

Observational measures of child development assess progress across multiple domains, including language, cognitive development, interpersonal/self-management skills, and approaches to learning. Unlike the screeners discussed in Section 4.2, these tools are embedded in ongoing classroom practice and are designed to capture growth over time rather than a single point-in-time snapshot. Because they rely on structured teacher or caregiver judgment, they often prioritize assessment in natural learning environments and instructional usefulness over

standardization. This design makes them valuable for formative purposes, but it also creates challenges for producing reliable and comparable results when scores are aggregated for system-level reporting, as differences in how field-based observers interpret and score behaviors can affect results.

Teaching Strategies GOLD

Teaching Strategies GOLD (TS GOLD), developed by Teaching Strategies, LLC, is an ongoing student-level observational assessment system designed to document the development of children from birth through kindergarten across multiple domains, including social-emotional, physical, language, and cognitive development. Widely adopted by state pre-k and Head Start systems, it is typically used by teachers to track individual growth for instructional planning and by administrators to report aggregate child-level developmental data.⁴⁴

Evidence indicates that TS GOLD is well suited for formative assessment, capturing the skills it is designed to measure. Its scores are more limited, however, when used as a proxy for program quality or as a high-stakes predictor of later achievement.⁴⁵ A recurring concern is rater subjectivity, including “halo effects,” in which children are rated similarly across domains in ways that appear to reflect overall impressions rather than domain-specific skill evidence. Rater subjectivity also includes the “reference bias,” previously described for KEAs. In practice, the subjectivity can reduce the tool’s diagnostic precision and limit its usefulness for distinguishing differences in classroom environments or program quality.⁴⁶ Furthermore, research suggests that teacher ratings from TS GOLD often do not capture enough variation in children’s skills to

reliably predict later achievement, particularly when compared to direct assessments that measure similar things at kindergarten entry.⁴⁷

From an alignment perspective, TS GOLD is a national observational assessment system designed to align broadly with early learning guidelines, including frameworks like Arkansas's CDELS. Although TS GOLD is conceptually consistent with the CDELS as a standards framework, it is not specifically customized for Arkansas and is distinct from the CDELS Developmental Rating Scale, a separate Arkansas-specific instrument designed to assess children directly against the CDELS standards.⁴⁸ TS GOLD measures broad developmental progressions rather than the specific discrete skill mastery often emphasized in early elementary assessments (e.g., phonics decoding tasks). Because of this, there can be a "measurement mismatch:" a child can appear "ready," or developmentally on track, according to TS GOLD's observational framework while still showing gaps on more granular K–2 academic assessments. Consequently, policy guidance consistently emphasizes that while TS GOLD provides a valuable snapshot for instructional planning, aggregating its data to evaluate program efficacy requires caution due to these concerns.⁴⁹

While TS GOLD is marketed as curriculum-agnostic, not just to be used with Teaching Strategies' *The Creative Curriculum*, programs using different High Quality Instructional Materials (HQIM) may find the alignment between the tool's assessment prompts and their instructional materials less clear. Additionally, other curriculum developers (e.g., *Frog Street*) offer their own aligned assessment tools. This means that providers using curricula other than

The Creative Curriculum may need to weigh the use of TS GOLD against assessments that are more directly aligned with their instructional materials.

- **Alignment Function:** Captures multiple dimensions of development through ongoing documentation in natural instructional settings, supporting a whole-child, growth-oriented definition of readiness.
- **Limitation:** Teacher judgment and documentation protocols introduce subjectivity and halo effects, which can complicate both the interpretation of scores and comparisons across the state.

The Work Sampling System (WSS)

The Work Sampling System (WSS) is a curriculum-embedded child-level observational assessment used from preschool through the elementary grades. Like TS GOLD, WSS relies on structured observations, portfolios, and developmental checklists to document children's progress across domains in natural classroom contexts.

The WSS has well-studied validity record, that is, there has been a lot of work answering whether the WSS measures what it claims to measure and the interpretations of its results are accurate and appropriate for the intended purpose. In preschool contexts, the Work Sampling for Head Start adaptation consistent and moderately strong relationships with individually administered early reading and mathematics assessments. It also showed evidence of explaining achievement outcomes beyond demographic factors.⁵⁰ Together, this evidence suggests alignment between this tool and direct assessments administered at the same point in time.

In later grades, one study found positive relationships between first-grade teacher ratings and third-grade performance on statewide assessments.⁵¹ However, in this study, once student demographic characteristics were considered, teacher ratings did not explain large, additional differences in later achievement.

Because WSS also depends on teacher judgment, concerns about equity across diverse populations and score comparability across the state are central concerns. As with other tools that depend on teacher subjectivity, research suggests that these scores can introduce risks when they are used for evaluative purposes, including differences in how behaviors are interpreted across racial and ethnic groups.⁵² Like TS GOLD, these concerns support treating WSS as an instructional and developmental monitoring tool rather than an accountability metric.

Overall, the evidence suggests that WSS captures meaningful information about children's development, particularly in early literacy and mathematics. At the same time, its unique contribution to predicting later high-stakes test performance is limited. This pattern is common across observational systems: teacher ratings align with children's skills in the moment, but they are not designed to function as precise forecasting tools for later accountability outcomes.

- **Alignment Function:** Aligns with whole child developmental philosophy. Uses classroom-embedded measurement by documenting progress across domains through authentic work products and structured teacher observation.

- **Limitation:** Teacher subjectivity, documentation protocol, and potential demographic differences in children may make interpretation of scores and comparisons across the state difficult.

The Ounce Scale

The Ounce Scale is an authentic, child-level observational assessment designed for infants and toddlers (birth to 42 months). It emphasizes developmental documentation in natural contexts and incorporates observational records, family input, and developmental profiles.

In a multi-site Early Head Start study of 287 infants and toddlers, the Ounce Scale produced generally dependable results across age groups, indicating that the tool captures children's development in a consistent way, though this was weaker for some of the youngest children.⁵³ When compared to standardized direct assessments administered at the same time, Ounce scores were meaningfully related to children's cognitive, language, and social-emotional performance. The strongest relationships were observed for social-emotional development.⁵⁴ Importantly, the study also found that Ounce scores helped identify children who may be at risk and provided information about their development that goes beyond what can be understood from basic child and family characteristics alone.⁵⁵

At the same time, the authors position the Ounce Scale explicitly as a performance-based assessment embedded in ongoing caregiving routines rather than as a norm-referenced achievement test or a long-term academic predictor.⁵⁶ Its evidence base supports its use for developmental monitoring and individualized planning rather than for high-stakes program evaluation.

- **Alignment Function:** The Ounce Scale supports ongoing monitoring of infant and toddler development within everyday caregiving routines. It helps teachers and families document growth over time and strengthens family engagement by making development visible in familiar, relationship-based settings.
- **Limitation:** The tool was not designed to predict later academic performance or to compare programs across sites. Because it relies on ongoing observation to collect evidence over time, its scores are not well suited for high-stakes accountability or statewide performance comparisons. Additionally, since the Ounce Scale relies on caregiver observation and documentation, the Ounce Scale is also subject to variation in how these adults interpret and record children’s behaviors, which can affect score comparability across settings.

Strength of Evidence and Appropriate Use
Alignment with later child outcomes.

Across tools, observational child development measures tend to show meaningful alignment with standardized assessments administered at the same point in time. Evidence is more limited for predicting later learning gains. Once differences in children’s background characteristics are taken into account, predictive relationships are typically smaller but can still be meaningful. Because these tools rely on teacher and caregiver judgment, they are better suited for documenting growth than for forecasting long-term outcomes.

Summary judgment.

Student-level observational systems such as TS GOLD, WSS, and the Ounce Scale are helpful tools for tracking children’s development within everyday classroom and caregiving routines. They support a whole-child view of growth and are valuable for instructional planning. At the same time, subjectivity in scores and limited predictive power of longer-term child outcomes limit their use as high-stakes accountability metrics or as standalone indicators of program quality.

Appropriate policy use based on evidence.

This group of tools is best aligned with instructional planning, professional learning, family communication, and developmental monitoring across domains that direct screeners do not capture. Aggregate reporting can support descriptive system monitoring when implemented carefully. However, research cautions against using observational scores to evaluate program effectiveness or rank programs without additional safeguards and clear evidence that such uses are valid for diverse groups of students across different types of settings.

The broader measurement literature cautions against “wishful linking,” or assuming that observational tools designed for formative purposes can automatically be used for program evaluation without additional validation.⁵⁷ Scholars emphasize that assessment systems must be evaluated not only for how the instrument performs, but also for whether its scores are being used in ways that are supported by evidence.⁵⁸

Key limitations affecting alignment.

Rater Subjectivity (Reference Bias and Halo Effects): Teachers and caregivers may rate children similarly across different domains (e.g., giving high scores in math because a child has good social skills), which reduces the tool's diagnostic precision. Furthermore, because student-level observational systems rely on adult interpretation of behavior, differences in how behavior is perceived and rated can affect comparability across student groups.

Limited Predictive Validity: Child-level observational ratings generally do not predict longer-term academic outcomes as consistently as direct assessments of discrete skills. This can be because the former often do not capture fine-grained variation in skills.

Aggregation Validity: When scores are combined across classrooms or sites, they may not reliably distinguish differences in program quality.

Construct Mismatch: The broad developmental progressions used in observational tools (e.g., "demonstrates knowledge of print") do not always align neatly with the discrete, constrained skills (e.g., "names all uppercase letters") emphasized in K–2 accountability tests. This can produce discrepancies across measures even when both are functioning as designed.

Global Environmental and Structural Quality Indicators

Beyond specific adult-child interactions (discussed in Section 4.1), the Arkansas QRIS relies on a set of "structural" indicators (i.e. *inputs*), such as staff qualifications, ratios, and administrative capacity, to create the necessary conditions for learning. Three of these elements have been called the "Iron Triangle" of quality in some research: group size, teacher-child ratio, and staff qualifications.⁵⁹ They are relatively easy to regulate and monitor. However, their

relationship to child outcomes is indirect: they appear to matter primarily because they enable higher-quality daily classroom interactions, rather than affecting children's learning on their own.

Teacher Qualifications and Experience

State quality systems frequently prioritize bachelor's degrees and specialized credentials for lead teachers, operating on the assumption that higher education leads to stronger instructional practices. However, empirical evidence suggests that formal degree attainment alone shows an inconsistent association with children's developmental outcomes.

Degrees and credentials.

Meta-analytic evidence indicates that higher levels of staff education are modestly associated with higher scores on global measures of early childhood environment quality, including observational tools such as the Early Childhood Environment Rating Scale (ECERS).⁶⁰ However, when studies attempt to isolate the effects of specific degree thresholds (e.g., bachelor's vs. associate's vs. high school), findings are more inconsistent, with some large-scale analyses showing that ECE teacher education and major were not consistently related to either classroom quality or children's academic gains.⁶¹

Other research focusing on child outcomes similarly concludes that associations between staff education and children's language, math, or social-emotional development are generally weak, null, or limited to particular domains such as language.⁶² In contrast, evidence from home-based early care settings suggests that associate's and bachelor's degrees, as well as specialized

early childhood coursework, may be associated with stronger global caregiving quality or more responsive adult–child interactions, though links to toddlers’ cognitive and social-emotional outcomes are limited and context dependent.⁶³

It is worth noting that the bulk of this research focuses on bachelor’s degree attainment and may not be consistent for ECE staff holding associate’s degrees or Child Development Associate (CDA) credentials, both of which are common among Arkansas’s ECE workforce. Taken together, this literature indicates that staff qualifications are more consistently related to the quality of the learning environment than to direct measures of children’s developmental outcomes.

Professional development.

Current evidence suggests that in-service professional development (PD) can meaningfully improve classroom quality, with coaching emerging as a particularly effective delivery format. A meta-analysis of 36 studies found a moderate to large overall effect of in-service training on process quality, with programs using coaching as the sole delivery format showing notably larger effects than other approaches.⁶⁴ A CLASS-focused meta-analysis found that coaching-based professional development produced moderate improvements in classroom interaction quality,⁶⁵ while a broader review found that individualized coaching outperformed one-off workshops, with the latter showing small or null effects on practice.⁶⁶

Consistent with these findings, a meta-analysis of language and literacy PD found that coaching intensity, but not more formal coursework intensity, predicted gains in process quality.⁶⁷ Effects of PD on child outcomes are consistently smaller than effects on classroom

quality but remain positive, and are strongest when child outcome measures align with the content focus of the PD program.⁶⁸ Neither curriculum implementation nor scale-based training content has been found to be more effective than other approaches, suggesting that how PD is delivered may matter more than its specific content focus.⁶⁹

Staff experience.

Research suggests that years of teaching experience is an inconsistent predictor of classroom quality in early childhood settings. A large study found that educator experience was largely unrelated to high-quality processes in classrooms after accounting for other structural characteristics (e.g., group size), and a systematic review similarly found few associations between educators' years of experience and children's outcomes.⁷⁰ This does not mean that experience lacks value, as more experienced educators may support program functioning in ways not captured by standard quality measures, such as mentoring staff or building long-term relationships with families. However, there is limited evidence that years of experience alone are directly associated with higher classroom quality or improved child outcomes.

Class Size and Child-Teacher Ratios

Low child–teacher ratios and small group sizes are widely recognized as foundational safety and structural quality standards. Recent meta-analytic evidence confirms a "threshold" theory, finding that ratios must be sufficiently low to *allow* for responsive interactions to occur. Specifically, one study identified non-linear thresholds, finding that in classrooms of children aged 3 to 5, ratios of 7.5:1 or lower and class sizes below 15 were associated with moderate to large improvements for cognitive outcomes, whereas improvements in ratios above this threshold

yielded diminishing returns.⁷¹ Another systematic review found that within the range of what is typically permissible by licensing regulations, better ratios in preschool-aged classrooms are not associated with better child outcomes,⁷² while a recent meta-analysis reported that there is no empirical evidence about ratios and child outcomes for children 2 years old or younger.⁷³

Like other structural variables, ratios and group size function as enabling conditions. They can make high-quality interactions more feasible, but do not guarantee them. Group size and child-to-adult ratio are the structural features most consistently linked to process quality, shaping what teachers do (e.g., more talk with children, more time in instruction, less time in transitions) though not necessarily how well they do it.⁷⁴ The only preschool class size experiment found that reducing classes from 20 to 15 produced more one-on-one interactions and greater literacy gains, but no differences in observed classroom interaction quality, which suggests that benefits may operate through increased individual contact.⁷⁵

Curriculum Implementation and Fidelity

While many state systems monitor the *presence* of an evidence-based curriculum as a structural indicator, research emphasizes that the *fidelity* of implementation is the primary driver of student outcomes. Simply having a published curriculum does not reliably improve child outcomes. Experimental evidence found that widely used whole-child curricula produced higher classroom quality scores than locally developed approaches but failed to improve children's school readiness while skill-targeted curricula did boost achievement in their targeted domains.⁷⁶ Across five large preschool samples, classrooms reporting use of a curriculum were largely indistinguishable from those without one, and classrooms using the same curriculum varied as

widely in quality as the broader population of classrooms.⁷⁷ Without ongoing professional development and monitoring of children's progress, curricular mandates may function more as structural checkboxes than as drivers of learning.

Unlike broad structural indicators, specific curricula focused on domain-specific skills (e.g., math or literacy) have shown moderate positive impacts on corresponding child outcomes when implemented with fidelity and supported by coaching on the specific instructional activities involved.⁷⁸ More generally, a recent book chapter reviews various ECE curriculum and any experimental evidence for each.⁷⁹

Global Environment & Administration (ERS & PAS)

In addition to staff metrics, states use observational tools to assess the physical and administrative capacity of programs.

Environment Rating Scales (ERS).

The Environment Rating Scales (ERS), including the *Early Childhood Environment Rating Scale* (ECERS) and the *Infant/Toddler Environment Rating Scale* (ITERS), and their home-based counterpart, the *Family Child Care Environment Rating Scale* (FCCERS), are observational instruments designed to assess the global quality of early learning environments by measuring a comprehensive set of indicators ranging from physical space and materials to personal care routines and teacher-child interactions (for a recent scoping review of 154 studies across 40+ countries, see Alpys & Hernández-Torrano, 2025).⁸⁰

Research evaluating the predictive validity, or the effectiveness of a tool in predicting future child developmental outcomes, of these scales indicates a consistently positive but small to modest associations, with relationships varying significantly by developmental domain. Some research demonstrates that higher global quality scores are statistically associated with gains in children's skills, particularly in language⁸¹ and social skills.⁸² However, ECERS-R scores were mostly unassociated with child outcomes assessed in studies,⁸³ and a validity analysis using nationally representative data suggests that inconsistent relationships may partly reflect limitations of the tool in consistently capturing different dimensions of the environmental quality.⁸⁴

Furthermore, scholarship supports a "threshold hypothesis," suggesting that ERS scores may only reliably predict positive child outcomes once a program exceeds a specific level of quality, typically in the "good" to "excellent" range, rather than showing a linear benefit at lower levels.⁸⁵ Validation studies of the newer ECERS-3 suggest its predictive power for individual school readiness remains limited compared to specific measures of interaction.⁸⁶

Program Administration Scale (PAS).

While the ERS measures the classroom environment, the Program Administration Scale (PAS) measures the quality of leadership and management practices in centers, evaluating dimensions such as human resources, fiscal management, and risk management, while the Business Administration Scale (BAS) assesses similar constructs in family child care homes. Empirical evaluations of the PAS indicate consistent associations between administrative quality and proximal program metrics, though connections to distal results remain unverified.

Studies have observed a statistically significant, positive relationship between PAS scores and measures of global classroom quality (ECERS/ITERS).⁸⁷ However, it has not been empirically linked directly to individual child developmental outcomes; current research supports the PAS/BAS primarily as a gauge of the organizational conditions that precede child learning rather than a direct predictor of child outcomes themselves.

Measures like the PAS/BAS focus on how well early childhood programs are led and managed. Strong leadership, stable staffing systems, and sound fiscal practices help programs operate effectively and create the conditions for high quality teaching. However, these organizational features do not directly produce learning gains for children. Instead, they make it more likely that teachers can provide consistent, engaging interactions that support development. Because this influence operates through the classroom experience, administrative and other structural indicators often show weaker statistical links to child outcomes, even though they remain essential building blocks of a strong early childhood system.

Strength of Evidence and Appropriate Use

Alignment with later child outcomes.

Research indicates that structural and global environmental indicators show indirect and sometimes inconsistent relationships with children’s developmental outcomes. Meta-analytic evidence suggests a nonlinear relationship between child–teacher ratios and cognitive outcomes for preschool-aged children. Meaningful gains tend to appear once ratios fall below approximately 7.5:1 and class sizes are relatively small.⁸⁸ However, a separate systematic review found limited evidence that within the range typically permitted by licensing standards, variation

in ratios is often not associated with measurable differences in child outcome suggesting the effect may be most consequential at the extremes.⁸⁹

Higher levels of teacher education are modestly associated with higher global environment ratings,⁹⁰ but degree attainment alone has not been a consistent predictor of classroom interaction quality or children's learning gains across large studies.⁹¹

Observational measures of environmental quality and administrative functioning capture features of program organization, leadership, and physical learning conditions. While some studies find small positive associations with child skill development, results are generally modest or inconsistent compared with measures of instructional interactions.

Summary judgment.

Taken together, these indicators primarily measure a program's capacity to support student learning rather than learning itself. They signal whether programs are safe, stable, and equipped with foundational resources, but they are limited proxies for the quality of instruction children actually experience.

Appropriate policy use based on evidence.

Structural standards establish the minimum conditions necessary for safe and manageable early learning environments and help monitor operational stability across center- and home-based settings. Furthermore, data on qualifications and staffing patterns can inform compensation policy, professional development strategies, and coaching investments that strengthen instructional practice.

Key limitations affecting alignment.

Indirect Pathway: Structural conditions influence children's development primarily through their effects on daily teaching interactions. Because this influence is mediated rather than direct, statistical associations with child outcomes are often modest.

Threshold Effects: Improvements in ratios or class size appear most consequential once programs reach relatively high-quality benchmarks (e.g., ratios below 7.5:1). Gains above or below these thresholds may produce limited observable impact.

Presence vs. Effective Use: Structural indicators capture whether resources or policies exist, not whether they are implemented in ways that meaningfully support instruction.

Early-Grade Academic and Developmental Outcomes as Alignment Anchors

To assess whether early childhood quality measures align with later outcomes, it is necessary to specify which outcomes serve as benchmarks for alignment. In this section, later outcomes refer to indicators of children's academic, social-emotional, and developmental functioning in kindergarten and the early elementary grades that are predictive of longer-term educational success.

Academic Outcomes

Early-grade academic outcomes, particularly in literacy and mathematics, are the most consistently studied indicators of educational trajectories. Empirical consensus confirms that school readiness skills are robust predictors of long-term attainment. Some large-scale analyses indicate that entry-level math and reading skills generally show moderate to large correlations with later elementary achievement, beyond looking at behavior skills alone.⁹²

However, the utility of these outcomes for alignment depends heavily on the *type* of skill measured. A key meta-analytic finding may help explain one part of the “fade-out” puzzle: McCormick et al. (2021) analyzed data from over 20,000 children and found that pre-K programs produced moderate initial impacts on constrained skills (e.g., letter naming), but these effects faded to near zero by kindergarten spring. In contrast, impacts on unconstrained skills (e.g., vocabulary) were smaller initially but showed no evidence of fade-out, suggesting they are more durable benchmarks for alignment.⁹³

Furthermore, recent evidence clarifies that alignment is bi-directional; the persistence of early gains is statistically dependent on whether K–2 instruction builds upon, rather than repeats, content mastered in pre-K.⁹⁴ Overall, the policy implication from this line of research is that alignment analyses that rely exclusively on constrained skill measures may systematically underestimate the long-term value of ECE programs that invest in vocabulary, self-regulation, and other unconstrained skills, which may include gains that are real and durable but not captured by spring kindergarten screeners.

Social-Emotional and Self-Regulation Outcomes

Social-emotional development and self-regulation are increasingly recognized as foundational for school success.⁹⁵ Skills such as attention, behavioral regulation, executive functioning, and peer relationships are associated with classroom engagement, learning behaviors, and longer-term outcomes including educational attainment and workforce participation.⁹⁶

While often treated as "soft skills," social-emotional and executive functioning (EF) capacities are robust empirical predictors of academic success. Research consistently identifies executive function as a moderate predictor of math and reading scores.⁹⁷ Longitudinal research indicates that learning related behaviors such as attention, persistence, and classroom engagement function as developmental mechanisms linking early classroom quality to later academic achievement.⁹⁸ On the other hand, higher quality instructional interactions are associated with stronger development of these behaviors, which in turn predict gains in children's math and reading skills over time.⁹⁹

Health and Developmental Indicators

Physical health, sensory-motor development, and approaches to learning are central components of whole-child frameworks, which emphasize that academic learning is supported by coordinated growth across cognitive, social-emotional, and physical domains.¹⁰⁰ However, these competencies are rarely incorporated into formal accountability systems focused on discrete academic skills.

Empirical evidence indicates that sensorimotor development is directly related to later academic performance. Longitudinal analyses show that fine motor skills at school entry predict subsequent gains in mathematics and reading achievement and explain unique variance in academic outcomes beyond early cognitive abilities.¹⁰¹ Similarly, children's opportunities to benefit from early education are shaped by health-related factors that influence attendance. Research on early chronic absenteeism demonstrates that preschool and early elementary absences are associated with weaker academic growth and increased likelihood of later learning

difficulties, reflecting reduced exposure to instructional opportunities during critical developmental periods.¹⁰²

Using early-grade outcomes as alignment anchors requires recognizing that no single outcome domain fully captures children's development. Because K–2 accountability systems tend to emphasize constrained academic skills, alignment analyses may underestimate the value of early childhood programs that produce gains in social-emotional functioning, engagement, or physical readiness but more modest impacts on discrete literacy measures.¹⁰³ In Arkansas, this dynamic contributes to a structural alignment challenge: the CDELS framework emphasizes unconstrained growth and motor/health development, while the ATLAS K–2 accountability system relies heavily on constrained academic screeners. Apparent misalignment may therefore reflect differences in what is measured rather than differences in program quality.

Strength of Evidence and Appropriate Use

Alignment with later child outcomes.

Early academic skills (math/reading) are the strongest predictors of later achievement, but "unconstrained" skills, including executive function, show more durable alignment than the "constrained" skills typically measured by K–2 screeners.

Summary judgment.

K–2 academic screeners provide necessary data on discrete skill mastery but are insufficient as the sole anchor for evaluating ECE program impact. Relying exclusively on

constrained academic skills to judge alignment likely underestimates the long-term value of high-quality early learning environments.

Appropriate policy use based on evidence.

Early-grade academic outcomes should be used as one of several indicators of how well early childhood and K–12 systems are aligned. Measures of learning-related behaviors, such as attention and self-regulation, should be considered alongside academic screeners to better reflect the full range of skills that predict children’s later success.

Key limitations affecting alignment.

Fade-out Artifacts: Misalignment could appear because K-2 assessments measure "constrained" skills (which may plateau) rather than "unconstrained" skills (which may persist).

Measurement Mismatch: ECE standards often emphasize "whole-child" domains, while K–2 assessments focus on narrow academic skills.

Data Gaps: There is a lack of systematic data collection on health, motor, and social-emotional domains in early elementary grades.

Academic Tilt: The heavy reliance on academic accountability in K–2 may obscure the benefits of ECE programs that prioritize process quality and social-emotional growth.

VI. Arkansas's Measurement Landscape

Policy Context in Arkansas

Arkansas is currently engaged in multiple efforts to strengthen early childhood quality measurement and alignment across birth-to-grade-2 systems. The revision of the state's QRIS, the implementation of the LEARNS Act, and the integration of the Office of Early Childhood within the Arkansas Department of Education have increased the focus on how early learning environments are measured and how those measures relate to later educational outcomes. These policy changes create an opportunity to clarify the conceptual framework linking early childhood quality indicators, kindergarten readiness measures, and early-grade academic outcomes.

Arkansas currently uses a range of tools related to early childhood quality and kindergarten readiness. These tools serve distinct purposes ranging from program improvement to accountability, yet discussions of these tools do not always clearly distinguish between

"readiness" and "quality." Table 1 summarizes these key tools, their intended purposes, and the evidence of their alignment with later outcomes.

Table 1. *Arkansas Measures and Evidence of Alignment*

Measure	Domain	Intended Purpose	Evidence of Alignment with Later Outcomes	Implementation Notes
CLASS®	Adult–child interactions	Program improvement, coaching	Moderate. Aligns with developmental mechanisms; Small to moderate, sometimes inconsistent associations with individual child outcomes.	Requires trained observers; limited score variation observed in many publicly funded settings; primarily designed for improvement rather than accountability.
Environment Rating Scales (ERS) & PAS/BAS	Global environment, materials, health/safety, & leadership	QRIS rating, technical assistance	Indirect. Consistently predicts quality of the <i>setting</i> with generally weak associations with individual child academic gains.	Metric used in current QRIS focuses on "capacity to educate" rather than instructional delivery.
KRIC	Developmental checklist	Parent and provider guidance, developmental monitoring	Limited. Few studies linking checklist scores to later child academic outcomes; results may be influenced by rater subjectivity.	Implementation varies; primarily a transition tool (family guidance) rather than system-level monitoring.
TS GOLD®	Observational child assessment	Instructional planning, formative assessment	Moderate. Population-level associations with later outcomes are	Implementation varies; reliability depends heavily on teacher training and

Measure	Domain	Intended Purpose	Evidence of Alignment with Later Outcomes	Implementation Notes
			adequate; individual child prediction is weaker due to rater subjectivity.	fidelity; not designed for high-stakes accountability.
ATLAS K Screener	Early literacy and mathematics assessment	Monitoring, instructional support, identifying for intervention	Strong for specific domains. High predictive validity for later reading/math scores; weak alignment with unconstrained/broad developmental domains.	Focused on constrained academic skills; not a comprehensive readiness indicator. Assessment chosen to meet LEARNS mandate. Limited data access for ECE providers.
Structural Indicators	Ratios, credentials, curriculum choice	Licensing, QRIS participation	Indirect. Weak direct associations with child outcomes; influence primarily via classroom process quality.	Policy-relevant system indicators; not designed to serve as direct proxies for child development or readiness.

Notes. CLASS® = Classroom Assessment Scoring System®; ERS = Environment Rating Scales (ECERS/ITERS/FCCERS); PAS/BAS = Program/Business Administration Scale; KRIC = Kindergarten Readiness Indicator Checklist; TS GOLD® = Teaching Strategies GOLD®; ATLAS = K-2 Arkansas Teaching, Learning & Assessment System Kindergarten Screener; QRIS = Quality Rating and Improvement System.

The strength and nature of alignment vary significantly across these measure types. First, measures of global capacity (ERS/PAS) and structural indicators confirm that a program is safe and organized enough to operate; however, these serve as "enabling conditions" for learning rather than direct drivers of student outcomes. Measures of classroom processes (e.g., CLASS®) capture the developmental mechanisms that drive growth (e.g., interactions), though they tend to show modest statistical relationships with longer-term child test scores. Conversely, early-grade

academic assessments (e.g., ATLAS) show stronger predictive relationships with later testing but capture narrower domains of development.

These distinctions imply that no single tool in Arkansas’s current portfolio is intended to be sufficient on its own. Rather, QRIS and early-grade policies should treat these measures as complementary pieces of a single alignment puzzle, that help us holistically understand ECE quality.

Mapping Arkansas Measures to Readiness Domains

To assess alignment, it is necessary to map each tool to the specific domains of development associated with kindergarten readiness. Research consistently identifies readiness as a multi-domain construct, including academic skills, social-emotional development, self-regulation, health, and approaches to learning. Table 2 maps Arkansas’s current measures against these domains.

Table 2. *Conceptual Alignment of Arkansas Measures with Kindergarten Readiness Domain*

Measure	Academic Skills (Literacy/Math)	Social-Emotional Development	Self-Regulation & Executive Function	Health & Physical Development	Approaches to Learning	Alignment Summary
CLASS®	Moderate	Moderate	Moderate	Not captured	Moderate	Captures classroom processes that support readiness-related mechanisms; modest to moderate direct prediction of child outcomes
ERS & PAS/BAS	Indirect	Indirect	Indirect	Direct/Strong	Indirect	The primary measure of physical environment, safety, and health; indirect link to learning.
KRIC	Limited (Checklist)	Broad Coverage (Subjective)	Broad Coverage (Subjective)	Broad Coverage (Subjective)	Broad Coverage (Subjective)	Broad checklist with limited validation and less well-established alignment with later outcomes; results may be influenced by rater subjectivity.
TS GOLD®	Moderate	Moderate	Moderate	Moderate	Moderate	Multi-domain observational tool; alignment depends on implementation fidelity and rater consistency.
ATLAS	Strong	Not captured	Not captured	Not captured	Not captured	Predictive of later academic outcomes with a narrower focus on academic domains; currently only for K-2.

Measure	Academic Skills (Literacy/Math)	Social-Emotional Development	Self-Regulation & Executive Function	Health & Physical Development	Approaches to Learning	Alignment Summary
Structural Indicators	Indirect	Indirect	Indirect	Not captured	Indirect	Enable quality conditions but are not directly related to child outcomes

Notes. CLASS® = Classroom Assessment Scoring System®; ERS = Environment Rating Scales (ECERS/ITERS/FCCERS); PAS/BAS = Program/Business Administration Scale; KRIC = Kindergarten Readiness Indicator Checklist; TS GOLD® = Teaching Strategies GOLD®; ATLAS = K-2 Arkansas Teaching, Learning & Assessment System. *Indirect* indicates that the measure influences the domain through learning environments rather than directly assessing child skills. *Not captured* indicates the domain is outside the tool’s design and measurement scope. *Broad Coverage (Subjective)* indicates the tool asks about the domain, but reliability is limited by rater judgment.

Table 2 shows a distinction in Arkansas’s educational measurement landscape. Specifically, the system emphasizes academic outcomes in early elementary grades yet prioritizes process quality in early childhood settings.

- Process measures (e.g., CLASS®) align most closely with social-emotional development and approaches to learning, or mechanisms that support long-term success but are not directly reflected in K–12 academic tests.
- Environment measures (e.g., ERS/PAS) provide the only standardized data on health and safety, domains that are foundational to development but distinct from K–12 academic monitoring.
- Academic screeners (e.g., ATLAS) provide strong signals for literacy and math trajectories but do not capture the broader "whole child" readiness domains emphasized in early childhood frameworks.

This mapping illustrates that the "misalignment" often cited in policy is partly a measurement artifact: the definition of success naturally shifts from *developmental support* in ECE to *academic proficiency* in K–12. From a policy perspective, this confirms that Arkansas’s current measures are functioning as designed for their specific aims, that is, program improvement in ECE transitioning to academic monitoring in K–12. Efforts to apply these distinct tools within a single predictive framework risk overlooking their primary value. These tools are complementary indicators of child well-being, rather than a single continuous metric of performance. At the same time, thoughtful choices on both sides of the ECE/K-12 transition will ultimately benefit children and families.

VII. Alignment and Gaps in Arkansas’s Current Approach

Areas of Alignment

Use of process quality measures aligned with developmental theory

Arkansas’s recent inclusion of CLASS® within the QRIS reflects a strong research base emphasizing supportive adult–child interactions as necessary to early learning and development. Measures of emotional support, classroom organization, and instructional support are theoretically aligned with the development of language and self-regulation that underpin later outcomes.

Integration of early-grade academic assessments

Arkansas’s new use of ATLAS in kindergarten through second grade aligns with evidence that early academic skills predict later academic outcomes, particularly when similar constructs are measured longitudinally. These assessments provide system-level information about early academic trajectories that can inform instruction and policy. One opportunity to strengthen this work involves improving access for ECE providers. Currently, ECE providers do not have access to ATLAS results for children previously enrolled in their programs, limiting the tool’s use as a feedback mechanism for ECE program improvement.

Multi-tool measurement system across early childhood and K–2

The presence of tools across birth-to-five and early elementary systems creates foundational infrastructure for alignment across early childhood and K–12. This multi-tool landscape provides opportunities for linking early learning experiences with later outcomes, as the state continues to develop a formal alignment framework.

Areas of Potential Misalignment or Gaps

Unclear Guidance on the Intended Purpose of Different Measures

Several tools used in Arkansas serve distinct purposes, including improvement, screening, monitoring, and accountability. In the absence of explicit guidance differentiating these purposes, there is risk that tools will be interpreted as interchangeable indicators of readiness or quality, even when they were not designed for such uses.

Use of CLASS® as a Proxy for Child Readiness or High-stakes Accountability

While CLASS® is theoretically aligned with mechanisms that support development, empirical evidence shows small to moderate, and sometimes inconsistent, relationships with individual child outcomes. Using CLASS® scores as a direct indicator of child readiness or as a high-stakes metric may extend beyond the tool's intended design and lead to interpretations that are not fully supported by the available evidence.

Variation in Implementation and Fidelity across Observational and Checklist-Based Tools

Observational measures and developmental checklists depend heavily on training, implementation fidelity, and rater judgment. Variation in implementation practices may affect comparability across settings and limit the interpretability of results for system-level decision-making. Ongoing calibration following initial training is important for maintaining consistency across these tools.

Narrow Domain Coverage of Academic Assessments

ATLAS and similar early-grade assessments focus primarily on academic domains. These tools do not capture social-emotional development, approaches to learning, executive

function, or health-related factors that research identifies as important for later success. Reliance on academic assessments alone may provide a more limited picture of readiness.

Development of a Formal Alignment Framework across Birth-to-Grade-2 Systems

Arkansas is in the process of developing a more unified framework clarifying how early childhood quality measures, kindergarten readiness indicators, and early-grade academic outcomes are intended to connect. Without such a framework, alignment is largely implicit, and stakeholders may draw conclusions about system coherence that are not empirically grounded.

Implications for QRIS Policy and Implementation

Arkansas's current QRIS incorporates multiple indicators intended to reflect program quality. These include structural requirements (e.g., ratios, staff credentials), global quality measures (i.e., ERS, PAS), and process requirements (e.g., curriculum use, child portfolios). As the state considers modifications to its QRIS, alignment with research-based indicators of kindergarten readiness could be a key policy priority.

The evidence reviewed in this report suggests that the current QRIS measures vary in the extent to which they align with later child academic outcomes:

- **Process Quality (CLASS®):** Measures of adult–child interactions align most closely with the developmental mechanisms (language, regulation) that support learning. While theoretically strong, direct statistical relationships with later academic assessments are typically small to modest, suggesting this type of tool supports developmental processes more than academic proficiency. However, this makes sense given the design of the tool to support overall classroom quality improvement.

- **Global Quality & Administration (ERS & PAS):** The Environment Rating Scales (ERS) and Program Administration Scale (PAS) effectively measure safety, organization, and material availability, key enabling conditions for program operation. However, research indicates these measures have weak and inconsistent relationships with individual child academic gains. Thus, these tools ensure a program is "ready to educate" but do not directly capture instructional effectiveness.
- **Structural Quality (Ratios & Credentials):** Arkansas tracks structural indicators such as teacher-child ratios, group sizes, and staff qualifications. While these are critical for safety and managing the classroom environment, research consistently shows they are "necessary but insufficient" drivers of child outcomes. They create the potential for quality interactions but do not directly measure them.
- **Child Assessments (or Portfolios):** Currently, Arkansas's QRIS requires programs to maintain child portfolios or conduct assessments. However, this is a compliance measure (verifying the activity occurs) rather than an outcome measure (tracking child growth).

These findings suggest that QRIS revisions should clarify the distinct role of each tool. Structural indicators and global quality measures (ERS/PAS) remain vital for establishing the safe, well-managed environments necessary for operation. CLASS[®] represents one of the most well-developed and widely used tools for capturing the quality of adult–child interactions, which research has identified as supporting early learning and development. Its content-agnostic design and applicability across age groups and program types make it well suited to serve as a common measure within a diverse early childhood landscape. Accordingly, CLASS[®] should be prioritized

as a component of QRIS for driving instructional improvement through coaching and professional development. Finally, child assessments can be more intentionally leveraged as a resource for teacher planning, without being used as a high-stakes program rating metric.

To support this clarification of purpose, Table 3 synthesizes the components of Arkansas’s current QRIS system. It maps each indicator to its alignment with research on child outcomes and identifies its most appropriate application within a policy framework.

Table 3. Conceptual Alignment of Current QRIS Measures with Later Outcomes

QRIS Indicator	Alignment with Later Academic Outcomes	Appropriate Policy Use
CLASS® (Interactions)	Theory-Based. Supports regulation/language; small to modest associations with academic gains.	Improvement: Coaching, PD, monitoring teacher-child interaction quality and targeting supports to programs with the greatest need.
ERS & PAS (Environment/Admin)	Indirect. Ensures safety/organization but limited associations with academic gains.	System Capacity: Licensing baseline, health/safety monitoring.
Structural (Credentials, Ratios)	Indirect. Necessary enabling conditions but limited associations with academic gains.	Foundational: Minimum standards to operate safely.
Curriculum & Assessment Use	Compliance-Based. Verifies use of instructional tools; does not directly capture child outcomes.	Instructional Support: Ensuring teachers have resources to plan and tools to monitor student learning and development.

Notes. CLASS® = Classroom Assessment Scoring System®; PD = Professional Development

Overall, the indicators from Arkansas’s QRIS shown in Table 3 include measures that align with research-based mechanisms linked to later outcomes, but the measures differ in their appropriate uses and evidentiary strength. Process quality measures, particularly CLASS®, are

best suited to support the workforce through targeted staff professional learning. Structural indicators and global quality measures (such as ERS/PAS) provide important information about system capacity and safety but are not designed to serve as direct indicators of child readiness or learning. Finally, child assessments provide useful information for instruction and early intervention but are not direct measures of program quality.

Clarifying these distinctions within QRIS policy guidance and implementation would improve the system's validity, reduce misinterpretation, and support more informed use of existing data for improvement and informed decision-making.

It is also worth acknowledging an important feature not currently present in Arkansas's measurement portfolio: no tool in the system is specifically designed to function as a provider-level performance accountability indicator, meaning a measure intended to differentiate programs based on their effectiveness at improving child-level outcomes. As emphasized above, early childhood education is a very rich and complex system. Because of this, it is difficult to summarize all the different measurement tools into a single quality index or score that is comparable across different ECE settings and populations served.

Arkansas's existing ECE quality measurement system is consistent with how state QRIS systems have historically been designed. Child care licensing regulations establish a regulatory floor, ensuring baseline health and safety standards for all providers. Arkansas's QRIS tiered rating and reimbursement structure provides an incentive for continuous quality improvement and a quality differentiation mechanism above that floor, a design that traces directly to the Race to the Top–Early Learning Challenge framework that catalyzed QRIS adoption nationally. Growth-based ECE provider accountability, or approaches that use child-level outcome data to

differentiate ECE programs in ways that account for the characteristics children bring with them to those programs, have not yet been developed, in Arkansas or broadly across the field. Such approaches are methodologically complex and remain an active area of research and development. Accordingly, this report has focused on the current measurement tools used in Arkansas and highlighted which domains each tool is particularly effective in supporting.

VIII. Recommendations Focused on Measurement and Alignment

As the analysis above makes clear, Arkansas's current measurement portfolio serves multiple complementary functions and is largely consistent with the design intent of state QRIS systems. The recommendations below focus on clarifying how existing tools should be used and interpreted.

The evidence reviewed in this report indicates that Arkansas possesses many of the necessary components for a coherent and useful early childhood measurement system. At the same time, clearer alignment and guidance can help ensure these measures are interpreted and used effectively. Differences in the purposes, strengths, and limitations of existing tools can create the potential for misinterpretation, particularly if measures designed for program improvement or screening are conflated with indicators of child readiness or program effectiveness.

The recommendations below focus on clarifying measurement purposes, strengthening alignment across birth-to-grade-2 systems, and supporting Arkansas's QRIS implementation in ways that are consistent with research and policy best practices.

1. **Clarify intended purposes for each measure used in QRIS.** Develop state guidance that explicitly distinguishes between tools intended for improvement (e.g., CLASS®),

system capacity (e.g., structural indicators, ERS/PAS), and instructional planning (e.g., child assessments). Clear rules of use help prevent misinterpreting improvement tools as high-stakes accountability metrics.

2. **Develop a cross-sector alignment framework across Birth-to-Grade-2.** Create a state framework that conceptually links QRIS, early child assessments, and K–2 outcomes. This framework should clarify how early childhood quality indicators function as the developmental counterpart to the structural and instructional standards already required in K–12 schools, helping stakeholders understand these systems as sequential stages within a unified cradle-to-career education continuum.
3. **Prioritize process quality measures for professional growth.** Use CLASS[®] and other observational tools primarily as drivers for coaching and professional development rather than primarily as point-in-time ratings. Since these tools capture the mechanisms of learning (interactions), they are most effective when used to give teachers specific feedback rather than to rank programs.
4. **Broaden the definition of readiness beyond academic metrics.** While ATLAS provides strong data on literacy and math, policy discussions regarding "readiness" should acknowledge the domains it excludes. Where feasible, complement academic data with indicators of social-emotional development and physical health to present a holistic picture of Arkansas students' needs.
5. **Maintain a stable set of indicators for longitudinal tracking.** To allow for meaningful trend analysis and informed decision-making, avoid frequent changes to the measurement portfolio. Establishing a consistent set of core indicators allows the state to track system-

level progress over time without the noise of constantly shifting metrics. At the same time, phasing in changes gradually over several years with time to collect feedback and adjust based on how measurement tools actually work in our state.

Ultimately, achieving alignment between early childhood quality and early grade outcomes does not require these distinct systems to use identical measures. Rather, it requires a coherent framework where the supportive environments of early childhood are understood as necessary precursors to the academic proficiency of elementary school.

By clarifying the unique purpose of each tool, ensuring that safety compliance, instructional improvement, and child assessment each play their specific role without extending tools beyond their intended purpose, Arkansas can move from a landscape of fragmented measures to a unified system. This approach preserves the developmental integrity of early childhood education while providing the actionable data needed to build a seamless pathway for student success from birth through second grade.

IX. End Notes

¹ While Arkansas's Child Development and Early Learning Standards (CDELS) provide a developmental framework for alignment beginning at birth, no psychometrically validated assessment tool currently exists for evaluating children against the CDELS, which limits its use as a practical alignment target at this time.

² Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early Childhood Assessment: Why, What, and How*. National Academies Press.

³ Farran, D. C. (2011). Rethinking School Readiness. *Exceptionality Education International*, 21(2), 5–15.

⁴ National Education Goals Panel. (1998). *Principles and Recommendations for Early Childhood Assessments* (L. Shepard, S. L. Kagan, & E. Wurtz, Eds.). Washington, DC: National Education Goals Panel.

⁵ Slutzky, C., & DeBruin-Parecki, A. (2019). *State-Level Perspectives on Kindergarten Readiness* (Research Report No. RR-19-07). Princeton, NJ: Educational Testing Service.

⁶ Arkansas Division of Child Care and Early Childhood Education. *Arkansas Kindergarten Readiness Indicators*. Little Rock, AR: Arkansas Department of Human Services.

⁷ Farran, D. C. (2011). Rethinking School Readiness. *Exceptionality Education International*, 21(2), 5–15.

⁸ Snow, K. L. (2006). Measuring School Readiness: Conceptual and Practical Considerations. *Early Education and Development*, 17(1), 7–41.

⁹ Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, 15(2), 147–166.; Zero to Three. (2023). *State of Babies Yearbook: 2023*. Washington, DC: Zero to Three.

¹⁰ Puccioni, J., Bruce, J., Patel, N., & Chamberlain, L. J. (2018). Parental Attitudes, Behaviors, and Barriers to School Readiness among Parents of Low-Income Latino Children. *International Journal of Environmental Research and Public Health*, 15(2), 188.

-
- ¹¹ Jun, J., Meissel, K., Cooper, M., & Rudd, G. (2025). A systematic review: Parental perspective on school readiness during the pre- and post-transition periods. *International Journal of Educational Research Open*, 9, 100486. <https://doi.org/10.1016/j.ijedro.2025.100486>.
- ¹² Bassok, D., Latham, S., & Rorem, A. (2016). Is Kindergarten the New First Grade? *AERA Open*, 2(1), 1–31.
- ¹³ Snow, K. L. (2006). Measuring School Readiness: Conceptual and Practical Considerations. *Early Education and Development*, 17(1), 7–41.
- ¹⁴ García, E. (2015). *Inequalities at the Starting Gate: Cognitive and Noncognitive Skills Gaps between 2010-2011 Kindergarten Classmates*. Washington, DC: Economic Policy Institute.
- ¹⁵ Regenstein, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). *Uses and Misuses of Kindergarten Readiness Assessment Results*. Chicago, IL: The Ounce of Prevention Fund (Start Early).
- ¹⁶ Bailey, D. H., Duncan, G. J., Odgers, C. L., & Yu, W. (2020). Persistence and fadeout of educational intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. <https://doi.org/10.1177/1529100620915848>; Farran, D. C. (2011). Rethinking School Readiness. *Exceptionality Education International*, 21(2), 5–15.
- ¹⁷ Regenstein, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). *Uses and Misuses of Kindergarten Readiness Assessment Results*. Chicago, IL: The Ounce of Prevention Fund (Start Early).
- ¹⁸ Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12(1), 3–9. <https://doi.org/10.1111/cdep.12260>; Miller-Bains, K., et al. (2017). "Examining the Validity of a Multidimensional Performance-Based Assessment." *AERA Open*; Pavelski-Pyle, R. (2002). "Best Practices in Assessing Kindergarten Readiness." *The California School Psychologist*.
- ¹⁹ Snow, K. L. (2006). Measuring School Readiness: Conceptual and Practical Considerations. *Early Education and Development*, 17(1), 7–41; Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, 40(2), 184–202.
- ²⁰ Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12(1), 3–9. <https://doi.org/10.1111/cdep.12260>; Hamre, B. K. (2014). Teachers' daily interactions with children: An essential
-

ingredient in effective early childhood programs. *Child Development Perspectives*, 8(4), 223–230.

<https://doi.org/10.1111/cdep.12090>; Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) manual: Pre-K*. Paul H. Brookes Publishing Company.; Vygotsky, L. S. (1978).

Interaction between learning and development. In M. Gauvain & M. Cole (Eds.), *Readings on the development of children* (pp. 34–40). Scientific American Books. (Original work published 1934).

²¹ Hamre, B. K. (2014). Teachers’ daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives*, 8(4), 223–230.

²² Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and child outcomes. *PLOS ONE*, 11(12), e0167660.

<https://doi.org/10.1371/journal.pone.0167660>; Wang, S., Liu, S., & Chang, H. (2024). Teacher-child interaction domains measured by the CLASS and children’s pre-literacy skills: A systematic review and meta-analysis. *Early Education and Development*, 35(3), 431–453. <https://doi.org/10.1080/10409289.2022.2159259>; Zheng, Q., Yin, H., & Wang, Y. (2023). A meta-analysis of teacher-child interaction and early child outcomes from the perspective of CLASS. *Early Child Development and Care*, 193(15–16), 1620–1651.

<https://doi.org/10.1080/03004430.2023.2273788>.

²³ Zheng, Q., Zheng, X., Liu, N., Wang, F., & Zhao, Y. (2023). A meta-analysis of teacher–child interaction and early child outcomes from the perspective of CLASS. *Early Child Development and Care*, 193(15–16), 1620–1651. <https://doi.org/10.1080/03004430.2023.2273788>.

²⁴ Zheng, Q., Yin, H., & Wang, Y. (2023). A meta-analysis of teacher-child interaction and early child outcomes from the perspective of CLASS. *Early Child Development and Care*, 193(15–16), 1620–1651. <https://doi.org/10.1080/03004430.2023.2273788>.

²⁵ Gordon, R. A., & Farran, D. (2022). Introduction to ECRQ special issue measuring quality in early care and education: Past, present, and future. *Early Childhood Research Quarterly*, 60, 374–378. <https://doi.org/10.1016/j.ecresq.2022.03.009>; Gordon, R. A., & Peng, F. (2020). Evidence regarding the domains of

the CLASS PreK in Head Start classrooms. *Early Childhood Research Quarterly*, 53, 23–39.

<https://doi.org/10.1016/j.ecresq.2020.01.008>.

²⁶ Finders, J. K., Duncan, R. J., Grimm, K. J., & Pianta, R. C. (2021). Variability in preschool CLASS scores and children's school readiness. *AERA Open*, 7, 23328584211038938. <https://doi.org/10.1177/23328584211038938>;

Gordon, R. A., & Peng, F. (2020). Evidence regarding the domains of the CLASS Pre-K in Head Start classrooms. *Early Childhood Research Quarterly*, 53, 23–39. <https://doi.org/10.1016/j.ecresq.2020.01.008>; Guerrero-Rosada, P.,

Pan, Y., McDonald Connor, C., & Phillips, D. A. (2021). Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study. *Early Childhood Research Quarterly*, 54, 1–12. <https://doi.org/10.1016/j.ecresq.2020.07.009>;

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, 28(2), 199–209.

<https://doi.org/10.1016/j.ecresq.2012.12.002>.

²⁷ Markowitz, A. J., Bassok, D., & Player, D. (2020). Simplifying quality rating systems in early childhood education. *Children and Youth Services Review*, 112, 104947. <https://doi.org/10.1016/j.chilyouth.2020.104947>.

²⁸ Kara Y., Farmer, D.E., Wright, A., Kamata, A., Hujar, J., and Baker, S.K. (2024). Examining the effects of prekindergarten to second grade classroom quality on early elementary achievement. *Frontiers in Education*, 9, 1370573. <https://doi.org/10.3389/feduc.2024.1370573>.

²⁹ Kara Y., Farmer, D.E., Wright, A., Kamata, A., Hujar, J., and Baker, S.K. (2024). Examining the effects of prekindergarten to second grade classroom quality on early elementary achievement. *Frontiers in Education*, 9, 1370573. <https://doi.org/10.3389/feduc.2024.1370573>

³⁰ Burchinal, M., Vandell, D. L., Belsky, J., Clarke-Stewart, A., & Owen, M. T. (2021). Relating early care and education quality to preschool outcomes: The same or different models for different outcomes? *Early Childhood Research Quarterly*, 55, 35–51. <https://doi.org/10.1016/j.ecresq.2020.10.005>; Hamre, B. K. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives*, 8(4), 223–230.

-
- ³¹ Snow, K. L. (2006). Measuring School Readiness: Conceptual and Practical Considerations. *Early Education and Development*, 17(1), 7–41.; Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). Classroom Assessment Scoring System (CLASS) manual: Pre-K. Paul H. Brookes Publishing Company.
- ³² Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12(1), 3–9. <https://doi.org/10.1111/cdep.12260>.; Burchinal, M., Vandell, D. L., Belsky, J., Clarke-Stewart, A., & Owen, M. T. (2021). Relating early care and education quality to preschool outcomes: The same or different models for different outcomes? *Early Childhood Research Quarterly*, 55, 35–51. <https://doi.org/10.1016/j.ecresq.2020.10.005>.; Goble, P., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2015). Gains in teacher–child interaction quality and children’s school readiness skills: Does coaching matter? *Early Childhood Research Quarterly*, 33, 114–124. <https://doi.org/10.1016/j.ecresq.2015.05.001>.; Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and child outcomes. *PLOS ONE*, 11(12), e0167660. <https://doi.org/10.1371/journal.pone.0167660>.; Wenz-Gross, M., Upshur, C., & Lecy, J. (2009). Pathways to kindergarten readiness: Evidence from a preschool intervention. *Journal of Applied Developmental Psychology*, 30(6), 738–754. <https://doi.org/10.1016/j.appdev.2009.05.001>.; Zheng, Q., Yin, H., & Wang, Y. (2023). A meta-analysis of teacher–child interaction and early child outcomes from the perspective of CLASS. *Early Child Development and Care*, 193(15–16), 1620–1651. <https://doi.org/10.1080/03004430.2023.2273788>.
- ³³ Snow, K. L. (2006). Measuring School Readiness: Conceptual and Practical Considerations. *Early Education and Development*, 17(1), 7–41.; Slutzky, C., & DeBruin-Parecki, A. (2019). State-Level Perspectives on Kindergarten Readiness (Research Report No. RR-19-07). Princeton, NJ: Educational Testing Service.
- ³⁴ Soland, J., & Sandilos, L. E. (2021). English learners and kindergarten readiness assessments: The role of teacher bias. *Educational Evaluation and Policy Analysis*, 43(1), 26–47.
- ³⁵ Arkansas Department of Education. (2004). Qualls Early Learning Inventory (QELI) Technical Report. ADE Division of Elementary and Secondary Education.
-

-
- ³⁶ Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early Childhood Assessment: Why, What, and How*. National Academies Press.; Soland, J., & Sandilos, L. E. (2021). English learners and kindergarten readiness assessments: The role of teacher bias. *Educational Evaluation and Policy Analysis*, 43(1), 26–47.
- ³⁷ Arkansas Better Beginnings. (2024). *Kindergarten Readiness Indicator Checklist*. Arkansas Department of Human Services.
- ³⁸ Soland, J., & Sandilos, L. E. (2021). English learners and kindergarten readiness assessments: The role of teacher bias. *Educational Evaluation and Policy Analysis*, 43(1), 26–47.
- ³⁹ Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428-1446.
- ⁴⁰ Hustedt, J. T., Jung, K., Friedman-Krauss, A. H., Barnett, W. S., & Slicker, G. (2021). *Impacts of the Arkansas Better Chance Program on Student Outcomes*. National Institute for Early Education Research.
- ⁴¹ Ackerman, D. J., & Barnett, W. S. (2005). *Prepared for kindergarten: What does "readiness" mean?* National Institute for Early Education Research.
- ⁴² Ackerman, D. J., & Barnett, W. S. (2005). *Prepared for kindergarten: What does "readiness" mean?* National Institute for Early Education Research.; Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early Childhood Assessment: Why, What, and How*. National Academies Press.
- ⁴³ Regenstien, E., Connors, M., Romero-Jurado, R., & Weiner, J. (2017). Uses and misuses of kindergarten readiness assessment results. *The Ounce: Policy Conversations*, 6.
- ⁴⁴ Friedman-Krauss, A. H., et al. (2024). *The State of Preschool 2023*. National Institute for Early Education Research.
- ⁴⁵ Lambert, R. G., Kim, D. H., & Burts, D. C. (2015). Construct validity of the Teaching Strategies GOLD assessment system: A confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 33(1), 1-17.
- ⁴⁶ Miller-Bains, K., Russo, J. M., Williford, A. P., DeCoster, J., & Cottone, E. A. (2017). Evaluation of the psychometric properties of Teaching Strategies GOLD. *Early Childhood Research Quarterly*, 41, 1-13.; Russo, J. M., Williford, A. P., Markowitz, A. J., Vitiello, V. E., & Bassok, D. (2019). Examining the validity of a widely used
-

school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly*, 48, 14–25.

⁴⁷ Bassok, D., & Markowitz, A. J. (2019). The validity of observational measures of quality. (University of Virginia EdPolicyWorks).

⁴⁸ For more information about the Arkansas CDELS Developmental Rating Scale see https://ecep.uark.edu/ar_cdels.php.

⁴⁹ Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early Childhood Assessment: Why, What, and How*. National Academies Press.

⁵⁰ Meisels, S. J., Xue, Y., & Shablott, M. (2008). Assessing Language, Literacy, and Mathematics Skills With Work Sampling for Head Start. *Early Education & Development*, 19(6), 963–981.
<https://doi.org/10.1080/10409280801971890>.

⁵¹ Gallant, D. J. (2009). Predictive validity evidence for an assessment program based on the Work Sampling System in mathematics and language and literacy. *Early Childhood Research Quarterly*, 24(2), 133–141.
<https://doi.org/10.1016/j.ecresq.2009.03.003>.

⁵² Gallant, D. J., & Moore, J. L., III. (2008). Ethnic-based equity in teacher judgment of student achievement on a language and literacy curriculum-embedded performance assessment for children in grade one. *Educational Foundations*, 22(1–2), 63–77.

⁵³ Meisels, S. J., Wen, X., & Beachy-Quick, K. (2010). Authentic Assessment for Infants and Toddlers: Exploring the Reliability and Validity of the Ounce Scale. *Applied Developmental Science*, 14(2), 55–71.
<https://doi.org/10.1080/10888691003697911>.

⁵⁴ Meisels, S. J., Wen, X., & Beachy-Quick, K. (2010). Authentic Assessment for Infants and Toddlers: Exploring the Reliability and Validity of the Ounce Scale. *Applied Developmental Science*, 14(2), 55–71.
<https://doi.org/10.1080/10888691003697911>.

-
- ⁵⁵ Meisels, S. J., Wen, X., & Beachy-Quick, K. (2010). Authentic Assessment for Infants and Toddlers: Exploring the Reliability and Validity of the Ounce Scale. *Applied Developmental Science*, 14(2), 55–71. <https://doi.org/10.1080/10888691003697911>.
- ⁵⁶ Meisels, S. J., Wen, X., & Beachy-Quick, K. (2010). Authentic Assessment for Infants and Toddlers: Exploring the Reliability and Validity of the Ounce Scale. *Applied Developmental Science*, 14(2), 55–71. <https://doi.org/10.1080/10888691003697911>.
- ⁵⁷ Macy, M., & Bagnato, S. J. (2024). Wishful LINKing: Authentic Environmental Measures Aligned with Programmatic Purposes. *Early Childhood Education Journal*, 52(2), 259–279. <https://doi.org/10.1007/s10643-022-01424-0>.
- ⁵⁸ Goldstein, J., & Flake, J. K. (2016). Towards a framework for the validation of early childhood assessment systems. *Educational Assessment, Evaluation and Accountability*, 28(3), 273–293. <https://doi.org/10.1007/s11092-015-9231-8>.
- ⁵⁹ Slot, P. (2018), “Structural characteristics and process quality in early childhood education and care: A literature review”, OECD Education Working Papers, No. 176, OECD Publishing, Paris, <https://doi.org/10.1787/edaf3793-en>.
- ⁶⁰ Manning, M., Wong, G. T. W., Fleming, C. M., & Garvis, S. (2019). Is Teacher Qualification Associated With the Quality of the Early Childhood Education and Care Environment? A Meta-Analytic Review. *Review of Educational Research*, 89(3), 370–415. <https://doi.org/10.3102/0034654319837540>.
- ⁶¹ Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry, G. T., Howes, C., Iriondo-Perez, J., Jeon, H. J., Mashburn, A. J., Peisner-Feinberg, E., Pianta, R. C., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: results from seven studies of preschool programs. *Child development*, 78(2), 558–580. <https://doi.org/10.1111/j.1467-8624.2007.01014.x>; Madill, R., Moodie, S., Zaslow, M., & Tout, K. (2015). Review of selected studies and professional standards related to the predictors of quality included in the national survey of early care and education (OPRE Report #2015-93b). *Child Trends*.
-

-
- ⁶² Falenchuk, O., Perlman, M., McMullen, E., Fletcher, B., & Shah, P. S. (2017). Education of staff in preschool aged classrooms in child care centers and child outcomes: A meta-analysis and systematic review. *PLOS ONE*, 12(8), e0183673. <https://doi.org/10.1371/journal.pone.0183673>.
- ⁶³ Schaack, D. D., Le, V. N., & Messan Setodji, C. (2017). Home-Based Child Care Provider Education and Specialized Training: Associations With Caregiving Quality and Toddler Social-Emotional and Cognitive Outcomes. *Early Education and Development*, 28(6), 655–668. <https://doi.org/10.1080/10409289.2017.1321927>.
- ⁶⁴ Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, 88(3), 401–433. <https://doi.org/10.3102/0034654317751918>.
- ⁶⁵ Egert, F., Dederer, V., & Fukkink, R. G. (2020). The impact of in-service professional development on the quality of teacher-child interactions in early education and care: A meta-analysis. *Educational Research Review*, 29, 100309. <https://doi.org/10.1016/j.edurev.2019.100309>.
- ⁶⁶ Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, 88(3), 401–433. <https://doi.org/10.3102/0034654317751918>.
- ⁶⁷ Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, 38, 97–115. <https://doi.org/10.1016/j.ecresq.2016.07.002>.
- ⁶⁸ Brunsek, A., Perlman, M., McMullen, E., Falenchuk, O., Fletcher, B., Nocita, G., Kamkar, N., & Shah, P. S. (2020). A meta-analysis and systematic review of the associations between professional development of early childhood educators and children’s outcomes. *Early Childhood Research Quarterly*, 53, 217–248. <https://doi.org/10.1016/j.ecresq.2020.03.003>.; Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, 88(3), 401–433. <https://doi.org/10.3102/0034654317751918>.
-

-
- ⁶⁹ Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of In-Service Professional Development Programs for Early Childhood Teachers on Quality Ratings and Child Outcomes: A Meta-Analysis. *Review of Educational Research*, 88(3), 401–433. <https://doi.org/10.3102/0034654317751918>.
- ⁷⁰ Hanno, E. C., Gonzalez, K. E., Jones, S. M., & Lesaux, N. K. (2021). Linking features of structural and process quality across the landscape of early education and care. *AERA Open*, 7, 23328584211044519.; McMullen, E., Perlman, M., Falenchuk, O., Kamkar, N., Fletcher, B., Brunsek, A., Nocita, G., & Shah, P. S. (2020). Is educators' years of experience in early childhood education and care settings associated with child outcomes? A systematic review and meta-analysis. *Early Childhood Research Quarterly*, 53, 171–184.
- ⁷¹ Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407–428.
- ⁷² Perlman, M., Fletcher, B., Falenchuk, O., Brunsek, A., McMullen, E., & Shah, P. S. (2017). Child-Staff Ratios in Early Childhood Education and Care Settings and Child Outcomes: A Systematic Review and Meta-Analysis. *PLOS ONE*, 12(1), e0170256. <https://doi.org/10.1371/journal.pone.0170256>.
- ⁷³ Dalgaard, N. T., Bondebjerg, A., Klokke, R., Viinholt, B. C., & Dietrichson, J. (2022). Adult/child ratio and group size in early childhood education or care to promote the development of children aged 0–5 years: A systematic review. *Campbell Systematic Reviews*, 18(2), e1239.
- ⁷⁴ Hanno, E. C., Gonzalez, K. E., Jones, S. M., & Lesaux, N. K. (2021). Linking features of structural and process quality across the landscape of early education and care. *AERA Open*, 7, 23328584211044519.; Slot, P. (2018). Structural characteristics and process quality in early childhood education and care: A literature review, OECD Education Working Papers, No. 176, OECD Publishing, Paris, <https://doi.org/10.1787/edaf3793-en>.
- ⁷⁵ Francis, J., & Barnett, W. S. (2019). Relating preschool class size to classroom quality and student achievement. *Early Childhood Research Quarterly*, 49, 49–58. <https://doi.org/10.1016/j.ecresq.2019.05.002>.
-

-
- ⁷⁶ Jenkins, J. M., Duncan, G. J., Auger, A., Bitler, M., Domina, T., & Burchinal, M. (2018). Boosting school readiness: Should preschool teachers target skills or the whole child? *Economics of Education Review*, 65, 107–125. <https://doi.org/10.1016/j.econedurev.2018.05.001>.
- ⁷⁷ Jenkins, J. M., Whitaker, A. A., Nguyen, T., & Yu, W. (2019). Distinctions Without a Difference? Preschool Curricula and Children’s Development. *Journal of Research on Educational Effectiveness*, 12(3), 514–549. <https://doi.org/10.1080/19345747.2019.1631420>.
- ⁷⁸ Nguyen, T., Jenkins, J. M., & Auger Whitaker, A. (2018). Are Content-Specific Curricula Differentially Effective in Head Start or State Prekindergarten Classrooms? *AERA Open*, 4(2), 233285841878428. <https://doi.org/10.1177/2332858418784283>.
- ⁷⁹ Engel, M., Attaway, D. S., & Denker, H. (2025). Curricula in early childhood education preschool settings. In L. Cohen-Vogel, P. Youngs, & J. Scott (Eds.), *Handbook of education policy research*. American Educational Research Association.
- ⁸⁰ Alpys, A., & Hernández-Torrano, D. (2025). Scoping Review of Early Childhood Environment Rating Scale (ECERS): Assessing and Validating Quality in Early Childhood Education Settings. *Early Education and Development*, 1–21. <https://doi.org/10.1080/10409289.2025.2585294>.
- ⁸¹ Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta-analysis. *PLOS ONE*, 12(6), e0178512. <https://doi.org/10.1371/journal.pone.0178512>; Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of Classroom Quality in Prekindergarten and Children’s Development of Academic, Language, and Social Skills. *Child Development*, 79(3), 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>; McDoniel, M. E., Townley-Flores, C., Sulik, M. J., & Obradović, J. (2022). Widely used measures of classroom quality are largely unrelated to preschool skill development. *Early Childhood Research Quarterly*, 59, 243–253.
-

-
- ⁸² Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta-analysis. *PLOS ONE*, 12(6), e0178512. <https://doi.org/10.1371/journal.pone.0178512>; McDoniel, M. E., Townley-Flores, C., Sulik, M. J., & Obradović, J. (2022). Widely used measures of classroom quality are largely unrelated to preschool skill development. *Early Childhood Research Quarterly*, 59, 243–253. <https://doi.org/10.1016/j.ecresq.2021.12.005>.
- ⁸³ Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta-analysis. *PLOS ONE*, 12(6), e0178512. <https://doi.org/10.1371/journal.pone.0178512>; Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D. M., & Howes, C. (2008). Measures of Classroom Quality in Prekindergarten and Children’s Development of Academic, Language, and Social Skills. *Child Development*, 79(3), 732–749. <https://doi.org/10.1111/j.1467-8624.2008.01154.x>; McDoniel, M. E., Townley-Flores, C., Sulik, M. J., & Obradović, J. (2022). Widely used measures of classroom quality are largely unrelated to preschool skill development. *Early Childhood Research Quarterly*, 59, 243–253.; Sabol, T. J., & Pianta, R. C. (2014). Do Standard Measures of Preschool Quality Used in Statewide Policy Predict School Readiness? *Education Finance and Policy*, 9(2), 116–164. https://doi.org/10.1162/EDFP_a_00127.
- ⁸⁴ Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology*, 49(1), 146–160. <https://doi.org/10.1037/a0027899>.
- ⁸⁵ Harding, J. F., Nguyen, T., & Aikens, N. (2025). Do Early Childhood Environment Rating Scale-3 and Pre-K Classroom Assessment Scoring System Scores Need to Reach Thresholds to Predict Children’s School Readiness? *Early Education and Development*, 36(8), 1952–1969. <https://doi.org/10.1080/10409289.2025.2535031>.
-

-
- ⁸⁶ Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the Early Childhood Environment Rating Scale – Third Edition (ECERS-3). *Early Childhood Research Quarterly*, 44, 242–256. <https://doi.org/10.1016/j.ecresq.2018.04.009>.
- ⁸⁷ Lower, J. K., & Cassidy, D. J. (2007). Child Care Work Environments: The Relationship With Learning Environments. *Journal of Research in Childhood Education*, 22(2), 189–204. <https://doi.org/10.1080/02568540709594621>; Yaya-Bryson, D., Scott-Little, C., Akman, B., & Cassidy, D. J. (2020). A Comparison of Early Childhood Classroom Environments and Program Administrative Quality in Turkey and North Carolina. *International Journal of Early Childhood*, 52(2), 233–248. <https://doi.org/10.1007/s13158-020-00268-2>.
- ⁸⁸ Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407–428.
- ⁸⁹ Perlman, M., Fletcher, B., Falenchuk, O., Brunsek, A., McMullen, E., & Shah, P. S. (2017). Child-Staff Ratios in Early Childhood Education and Care Settings and Child Outcomes: A Systematic Review and Meta-Analysis. *PLOS ONE*, 12(1), e0170256. <https://doi.org/10.1371/journal.pone.0170256>.
- ⁹⁰ Manning, M., Wong, G. T. W., Fleming, C. M., & Garvis, S. (2019). Is Teacher Qualification Associated With the Quality of the Early Childhood Education and Care Environment? A Meta-Analytic Review. *Review of Educational Research*, 89(3), 370–415. <https://doi.org/10.3102/0034654319837540>.
- ⁹¹ Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry, G. T., Howes, C., Iriondo-Perez, J., Jeon, H. J., Mashburn, A. J., Peisner-Feinberg, E., Pianta, R. C., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: results from seven studies of preschool programs. *Child development*, 78(2), 558–580. <https://doi.org/10.1111/j.1467-8624.2007.01014.x>; Madill, R., Moodie, S., Zaslow, M., & Tout, K. (2015). Review
-

of selected studies and professional standards related to the predictors of quality included in the national survey of early care and education (OPRE Report #2015-93b). Child Trends.

⁹² Claessens, A., & Engel, M. (2013). How Important is Where you Start? Early Mathematics Knowledge and Later School Success. *Teachers College Record: The Voice of Scholarship in Education*, 115(6), 1–29.

<https://doi.org/10.1177/016146811311500603>; Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446.
<https://doi.org/10.1037/0012-1649.43.6.1428>.

⁹³ McCormick, M., Weiland, C., Hsueh, J., Pralica, M., Weissman, A. K., Moffett, L., Snow, C., & Sachs, J. (2021). Is Skill Type the Key to the PreK Fadeout Puzzle? Differential Associations Between Enrollment in PreK and Constrained and Unconstrained Skills Across Kindergarten. *Child Development*, 92(4), e599–e620.
<https://doi.org/10.1111/cdev.13520>.

⁹⁴ Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and Fade-Out of Educational-Intervention Effects: Mechanisms and Potential Solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. <https://doi.org/10.1177/1529100620915848>.

⁹⁵ McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, 43(4), 947–959. <https://doi.org/10.1037/0012-1649.43.4.947>.

⁹⁶ Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early Social-Emotional Functioning and Public Health: The Relationship Between Kindergarten Social Competence and Future Wellness. *American Journal of Public Health*, 105(11), 2283–2290. <https://doi.org/10.2105/AJPH.2015.302630>; Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698. <https://doi.org/10.1073/pnas.1010076108>.

⁹⁷ Devlin, B. L., Ellis, A., Zehner, T. M., Duncan, R. J., Elicker, J., Purpura, D. J., & Schmitt, S. A. (2024).

Contributions of preschool behavioral self-regulation and social skills to growth in different domains of early math knowledge. *Journal of Experimental Child Psychology*, 241, 105867. <https://doi.org/10.1016/j.jecp.2024.105867>;

Ernst, J. R., Grenell, A., & Carlson, S. M. (2022). Associations between executive function and early math and literacy skills in preschool children. *International Journal of Educational Research Open*, 3, 100201.

<https://doi.org/10.1016/j.ijedro.2022.100201>; Ruffini, C., Berni, M., Pierucci, G., & Pecini, C. (2024). Executive functions as predictors of learning prerequisites in preschool: A longitudinal study. *Trends in Neuroscience and Education*, 36, 100239. <https://doi.org/10.1016/j.tine.2024.100239>;

⁹⁸ Schmerse, D. (2020). Preschool Quality Effects on Learning Behavior and Later Achievement in Germany: Moderation by Socioeconomic Status. *Child Development*, 91(6), 2237–2254. <https://doi.org/10.1111/cdev.13357>.

⁹⁹ McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, 21(4), 471–490.

<https://doi.org/10.1016/j.ecresq.2006.09.003>; Nesbitt, K. T., Farran, D. C., & Fuhs, M. W. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology*, 51(7), 865–878. <https://doi.org/10.1037/dev0000021>.

¹⁰⁰ Bai, P., Johnson, S., Trost, S. G., Lester, L., Nathan, A., & Christian, H. (2021). The Relationship between Physical Activity, Self-Regulation and Cognitive School Readiness in Preschool Children. *International Journal of Environmental Research and Public Health*, 18(22), 11797. <https://doi.org/10.3390/ijerph182211797>; Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>.

¹⁰¹ Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine Motor Skills and Executive Function Both Contribute to Kindergarten Achievement. *Child Development*, 83(4), 1229–1244. <https://doi.org/10.1111/j.1467-8624.2012.01768.x>; Dinehart, L., & Manfra, L. (2013). Associations Between Low-Income Children's Fine Motor Skills in Preschool and Academic Performance

in Second Grade. *Early Education & Development*, 24(2), 138–161.

<https://doi.org/10.1080/10409289.2011.636729>; Grissmer, D., Grimm, K. J., Aiyer, S. M., Murrah, W. M., & Steele, J. S. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. *Developmental Psychology*, 46(5), 1008–1017. <https://doi.org/10.1037/a0020104>.

¹⁰² Connolly, F., & Olson, L. S. (2012). Early elementary performance and attendance in Baltimore City schools' pre-kindergarten and kindergarten. Baltimore Education Research Consortium.

<https://files.eric.ed.gov/fulltext/ED535768.pdf>; Ehrlich, S. B., Gwynne, J. A., & Allensworth, E. M. (2018). Pre-kindergarten attendance matters: Early chronic absence patterns and relationships to learning outcomes. *Early Childhood Research Quarterly*, 44, 136–151. <https://doi.org/10.1016/j.ecresq.2018.02.012>.

¹⁰³ McCormick, M., Weiland, C., Hsueh, J., Pralica, M., Weissman, A. K., Moffett, L., Snow, C., & Sachs, J. (2021). Is Skill Type the Key to the PreK Fadeout Puzzle? Differential Associations Between Enrollment in PreK and Constrained and Unconstrained Skills Across Kindergarten. *Child Development*, 92(4), e599–e620. <https://doi.org/10.1111/cdev.13520>.