

# Review of Early Childhood Assessments

## Methods and Recommendations

April 2017



## **Report Authors**

Annie Wright

Director of Evaluation

Dylan Farmer

Project Manager

Scott Baker

Research Professor

# Table of Contents

<b>EXECUTIVE SUMMARY .....</b>	<b>4</b>
<b>METHOD.....</b>	<b>4</b>
THE PANEL.....	4
THE PROCESS .....	5
<i>Background &amp; Planning</i> .....	5
<i>Meeting 1</i> .....	6
<i>Meeting 2</i> .....	8
<i>Meeting 3</i> .....	9
<b>WHAT WAS REVIEWED.....</b>	<b>10</b>
<b>REVIEW APPROACH.....</b>	<b>11</b>
CONTENT.....	11
PSYCHOMETRICS .....	14
ADMINISTRATIVE FEASIBILITY .....	17
PLACEMENT ON LISTS.....	17
<b>RECOMMENDATIONS FOR USE .....</b>	<b>19</b>
GENERAL RECOMMENDATIONS.....	19
COMBINING RECOMMENDED ASSESSMENTS TO MEASURE MULTIPLE DIMENSIONS.....	19
<i>Intended Purpose</i> .....	20
<i>Administering Tests Correctly</i> .....	20
<i>Interpreting Assessment Results</i> .....	20
<i>Avoiding the Misuse of Assessment Results</i> .....	21
<b>REFERENCES.....</b>	<b>22</b>
<b>APPENDICES .....</b>	<b>24</b>
A: LIST OF PANELISTS .....	24
B: STANDARDS ALIGNMENT .....	24
C: STATE OF ASSESSMENTS – PRE-K AND KINDERGARTEN .....	24
D: REQUESTS FOR INFORMATION (RFIs) – PRE-K AND KINDERGARTEN .....	24
E: RUBRIC AND GLOSSARY.....	24
F: MASTER SCORING SHEET .....	24
G: TOOL FOR REVIEWING ADMINISTRATIVE FEASIBILITY .....	24
H: FINAL LIST OF RECOMMENDED PRE-K ASSESSMENTS .....	24
I: FINAL LIST OF RECOMMENDED KINDERGARTEN ASSESSMENTS.....	24

# Executive Summary

---

*CORE convened an expert panel to review Prekindergarten progress monitoring and kindergarten readiness assessments and to make recommendations for use.*

The Center on Research and Evaluation (CORE) at Southern Methodist University was awarded Texas Education Agency (TEA) RFP #701\_17\_018 to convene an expert panel to review prekindergarten assessments for monitoring student progress (i.e., progress monitoring assessments) and kindergarten assessments for helping to determine kindergarten readiness (also referred to as kindergarten screening tools or screeners). The panel reviewed submitted assessments according to standard psychometric and measurement properties, assessment content, and features related to assessment usability and practicality in public school settings. Overall numerical scores for (a) assessment psychometrics and (b) assessment content coverage were calculated for each assessment. Then an overall weighted score that reflected each assessment's psychometric score and content coverage score was calculated. Assessments were ranked by this single overall score and are presented in two lists: (1) recommended kindergarten assessments and (2) recommended pre-kindergarten assessments.

This document summarizes the methodology CORE used to conduct the review process and presents the final lists of recommended assessments.

## Method

---

*CORE convened an expert panel, hosted three in-person meetings, trained reviewers on scoring the assessments, and summarized the rating scores to generate the lists of recommended assessments.*

### The Panel

Fourteen experts representing multiple disciplines in early childhood practice and research served on the early child assessment review panel. Panelists had expertise across a range of content areas (e.g., math, literacy, language, social & emotional learning). Panelists also included individuals with methodology and psychometric expertise, assessment design and development expertise, preschool and elementary school administrators and leaders (including traditional public and charter), researchers, community providers, and classroom teachers. See Appendix A – List of Panelists.

CORE project staff led the work to convene the panel, develop scoring systems, solicit ratings, and summarize and synthesize scores into the final recommended lists.

CORE used the following definitions to operationalize the terms “screeener” and “progress monitor”:

A **screening tool** involves brief assessments that are reliable, valid, and to the greatest extent possible, based on evidence of their use in high-quality studies and evaluations. They are conducted with all students or with targeted groups of students to identify those who are at risk of difficulty in specific areas and, therefore, need additional or alternative forms of instruction or intervention to supplement the instruction typically provided.

A **progress monitoring tool typically** involves brief assessments that are reliable, valid, and to the greatest extent possible, based on evidence of their use in high-quality studies and evaluations. They are conducted regularly with students (2–3 times per year, minimum, and as frequently as every other week or monthly) to determine the progress a student is making over time. Progress is frequently calculated as a slope of improvement score based on two or more administrations of the measure. This type of slope estimate is used as the basis for determining the adequacy of student progress, typically in relation to external progress criteria.

Some kindergarten readiness assessments were submitted as one-time screeners only, but the majority were submitted as instruments that could also be used as progress monitoring assessments. In most cases, the first assessment given at the beginning of the kindergarten year being appropriately used as an initial screener, and meant to be followed up with subsequent administrations to monitor progress.

## The Process

### Background & Planning

Starting in January 2017, CORE project staff began background research to anchor the review and rating process. This included two broad areas: (1) a review of assessment strategies and content areas used by other aspirant or exemplar states, and (2) review of standard scientific processes for executing a rigorous and effective review of education based assessment tools. (See References).

### **Background on Content**

The investigation into other states’ approaches to identify early childhood assessment indicated that Texas’s approach to standards was also seen elsewhere. Standards in the areas of early literacy, language and communication and mathematics are highly aligned across states, while there is some variability in how states emphasize and operationalize standards related to health, social development and general cognition. An extensive review of early learning policy and standards analyses lead CORE to systematically align Texas early learning standards with those of several other key states that are noted for robust and comprehensive standards. In addition to reviewing standards, a 50-state review of early learning assessment standards and polices was reviewed and compiled to determine common best-practice approaches to wide-scale evaluation and assessment of prekindergarten and kindergarten entry growth and proficiency. See Appendix B – Standards Alignment and Appendix C – State of Assessments.

### **Background on Psychometrics**

In addition to CORE staff’s assessment expertise, additional external sources were used to a compile an initial list of the psychometric features on which the panel would rate each assessment (see References). The sources utilized provided conceptual frameworks, examples of actual tools that have been used to evaluate the quality of educational assessments, which typically also included an overview and rationale for coding assessment tools and/or actual instruments used to rate student assessment instruments<sup>1</sup>. This information was compiled and provided to panelists for review and feedback prior to and during Meeting 1.

---

<sup>1</sup> For example: Center on Response to Intervention website (American Institutes for Research) Response (<http://www.rti4success.org/resources/tools-charts>) and The National Center on Intensive Intervention (<http://www.intensiveintervention.org/chart/progress-monitoring>)

## Meeting 1

The first panel meeting was held February 15, 2017, on SMU's Dallas campus.

Prior to the meeting, panelists received a synthesis document prepping them for the meeting and summarizing the content areas included in the Texas prekindergarten and kindergarten standards, and summaries of the content areas that were also aligned to standards in other key states. Panelists also received the Request for Information that TEA posted so that they would understand the types of information they would receive for review. See Appendix D – RFIs for Kindergarten and Prekindergarten assessments.

Meeting 1 focused on providing an overview of the project and completing two review related tasks.

For Task 1, panelists reviewed **content** included in Texas's five main domains and underlying concepts. Panelists provided input about additional concepts that could be added within each domain to be included in the review. For instance, the panel specified that letter and word writing should be included as an assessment area within Conventions in Writing (Emergent Literacy-Writing), and that sensory sensitivity should be included as an assessment area in Gross and Fine Motor concept (in the Health and Wellness Domain). CORE also solicited feedback from panelists about the comprehensiveness of the five domains, and the panelists agreed that it was not necessary to add additional domains for review.

For Task 2, panelists were asked to review a list of **psychometric features** that are common in scientific measurement, including educational assessments, to assess critical quality aspects of assessment tools (e.g., reliability, validity). Panelists' ratings of the overall importance of each psychometric feature for the objectives of the project provided CORE with initial information about what to include, and how to weight, the various psychometric features of the assessments.

### **Rubric & Glossary**

CORE compiled information from the literature review as well as feedback from panelists to develop a rubric (i.e., scoring system) for how assessment content and assessment psychometric features would be reviewed. The scores and rating scales are provided in a comprehensive glossary. See Appendix E – Rubric and Glossary.

### **Receipt of Assessments**

By 3:00pm on Friday March 17, 2017, CORE had received 19 prekindergarten progress monitoring assessments and 35 kindergarten readiness assessments. Some assessments were submitted as both pre-kindergarten progress monitoring and kindergarten readiness assessments, resulting in a list of 42 unique assessments for review. See Table 1 for the full listing of received and reviewed assessments.

**Table 1: Full List of Assessments Reviewed by Panel**

1. ABC Mouse First Grade Readiness	Kinder
2. ABC Mouse Kindergarten Readiness	Kinder
3. aimswebPlus	Kinder
4. BASC-3 BESS	Kinder & PK
5. Children's Progress Academic Assessment (CPAA) - Eng	Kinder & PK
6. Children's Progress Academic Assessment (CPAA) - Span	Kinder & PK
7. CIRCLE Progress Monitoring	PK
8. CIRCLE/C-PALLS+STEM (English) (electronic admin of CIRCLE)	PK
9. CIRCLE/C-PALLS+STEM (Spanish) (electronic admin of CIRCLE)	PK
10. Connect4Learning: Pre-K Assessment	PK
11. DESSA-mini	Kinder
12. Developmental Reading Assessment 2nd Edition PLUS (DRA2+)	Kinder
13. Devereux Early Childhood Assessment	PK
14. DIAL-4	Kinder & PK
15. DIBELS 6th Edition	Kinder
16. DIBELS Next Edition	Kinder
17. easyCBM	Kinder
18. El Inventario de Lectura en Espanol de Tejas (Tejas LEE)	Kinder
19. Evaluacion del desarrollo de la lectura, 2nd ed., plus K-6 (EDI.2+)	Kinder
20. Fountas & Pinnell Benchmark Assessment System K-2 (BAS)	Kinder
21. Frog Street Assessment	PK
22. GOLD	Kinder & PK
23. Indicadores Dinamicos del Exito en la Lectura (IDEL)	Kinder
24. Istation's Indicators of Progress -Early Reading (ISIP-ER)	Kinder & PK
25. Kinder Reading (K Ready) - English	Kinder
26. Kinder Reading (K Ready) - Spanish	Kinder
27. Learning Accomplishment Profile 3 (LAP-3)	PK
28. LION for Reading	Kinder
29. MAP for Primary Grades	Kinder
30. Oral Language Acquisition Inventory	Kinder & PK
31. PPVT	Kinder & PK
32. Preschool Language Scales (PLS-5)	Kinder & PK
33. Reading Inventory	Kinder
34. Ready, Set, K!	Kinder & PK
35. Sistema de evaluacion de la lectura (Sistema-SEL)	Kinder
36. Social Skills Improvement System (SSIS)	Kinder & PK
37. Spring Math	Kinder
38. Star Early Literacy	Kinder
39. Star Early Literacy Spanish	Kinder
40. Texas Kindergarten Entry Assessment (TX-KEA)	Kinder
41. Texas Primary Reading Inventory (TPRI)	Kinder
42. Work Sampling System	Kinder & PK

## **Assignment of Assessments to Reviewers**

CORE assigned panelists to specific assessments for review. To the degree possible, two panelists scored each assessment independently, and three review teams were created to review assessment content and assessment psychometrics. One team was comprised of individuals with sufficient content and psychometric knowledge to review their assigned assessments in both areas. Each member of this team reviewed between 4 and 5 assessments. A second team of educational practitioners and administrators reviewed their assigned assessments for content only. A third team reviewed just the psychometrics of those assessments reviewed by the second team. The second and third teams were assigned to 7 or 8 assessments.

## **Development of Scoring Tools; Review Matrices**

CORE developed two separate scoring tools, which are referred to as review matrices: (1) Review Matrix for Kindergarten Screeners and Progress Monitors and (2) Review Matrix for Prekindergarten Progress Monitors. These excel sheets served as data entry forms for panelists to use when reviewing their assigned assessments. They were pre-filled with all content areas and psychometric features that would be reviewed for screeners and/or for progress monitoring assessments. Drop down menus in the matrices were based on the scoring rubric in the glossary. The matrices also contained formulas that weighted scores based on panelists' feedback on what aspects of the assessments should be considered most important in review process. (Later sections of this report describe the scoring weights in detail).

## **Meeting 2**

At Meeting 2, panelists first reviewed the list of received assessments (Table 1) and indicated any conflicts of interest. In four cases, initial panelist assignments were revised based on conflicts of interest

Then panelists were shown the review areas (Table 2), and reviewed the glossary and the review matrices. Panelists then participated in a group scoring process of a selected assessment and shown a “live” use of the matrices. During this process, consensus was established on scores (for example, whether a concept should receive a 0, a 1, or a 2, and why). Additional information was added to the glossary to reflect the group’s decisions, and a final “frequently asked questions” document was developed and shared.

The meeting concluded by each panelist receiving all of the review material they would need to review each of their assigned assessments.

## **Review Check-in**

During the review weeks, CORE held weekly office hours where panelists could call or come in with questions, or submit questions via email. Minor clarifications regarding review procedures were discussed but no substantive changes were made to the review requirement, protocols, or coding process. Panelists submitted their scores on the online system to CORE by Tuesday March 11, 2017.

## **Summary and Synthesis of Ratings**

CORE compiled the panelists' submitted review matrices for all assessments. Score discrepancies between the two reviews of each assessment were identified. Scores were within one point of each other (e.g., a 0 and a 1, or a 1 and a 2) were considered a “match” and were not resolved. Score discrepancies greater than one point (e.g., a 1 and a 3) were reconciled by CORE project staff. The reconciled scores were then logged in the master scoring sheet. See Appendix F – Master Scoring Sheet.

### Meeting 3

Meeting 3 included discussions about weighting content and psychometric scores in a manner that provided greater assurance that assessments with acceptable levels of psychometric evidence would be on the list of recommended assessments. CORE then developed a weighting procedure that placed more emphasis on the psychometric properties of the assessments than the coverage of content. (see later sections for detailed information about the scoring process).

Meeting 3 also included discussion about assessment features related to the feasibility and utility of administration. Administrative features, including how efficiently assessments could be administered in real-world settings, ranged from issues related to cultural relevance, teacher and student use, language, time to administer, training needs and other pragmatic issues. See Appendix G – Administrative Features Tool. Recommended assessments were reviewed for six categories of administrative feasibility: (1) teacher friendly, (2) administrator friendly, (3) administrative features, (4) language, (5) cultural competence, and (6) student friendly.

# What Was Reviewed

Review areas fell into three main categories:

- (1) Administrative/feasibility
- (2) Content coverage (content validity) across five domains
- (3) Psychometric features

See Table 2 below, for a summary of concepts that were reviewed for each assessment and Appendix E for a **glossary** of how each of the areas were rated by panelists.

**Table 2: Areas Reviewed by Panel**

Administrative/Feasibility	Content	Psychometric
<p>Logged but not scored:</p> <ul style="list-style-type: none"> <li>• Title</li> <li>• Publisher</li> <li>• Recommended Use</li> <li>• Price per Student</li> <li>• Format: Direct or Obs</li> <li>• Format: Group or 1-on-1</li> <li>• Scoring formats</li> <li>• Language</li> <li>• Grade Levels</li> <li>• Test format (e.g., pp)</li> <li>• Requirements to admin</li> <li>• Time requirement per student</li> </ul> <p>Rated for informative purposes, not used in scoring:</p> <ul style="list-style-type: none"> <li>• Teacher friendly</li> <li>• Administrator friendly</li> <li>• Student Friendly</li> <li>• Administration format</li> <li>• Language</li> <li>• Cultural Relevance</li> </ul>	<p>Content areas rated by panelists and scored in order to make recommendations:</p> <ul style="list-style-type: none"> <li>• Emergent Literacy – Reading (5 constructs)</li> <li>• Emergent Literacy-Writing (3 constructs)</li> <li>• Language &amp; Communication (6 constructs)</li> <li>• Health &amp; Wellness (6 constructs)</li> <li>• Mathematics (5 constructs)</li> </ul>	<p>Psychometric features rated by panelists and scored in order to make recommendations:</p> <ul style="list-style-type: none"> <li>• Reliability</li> <li>• Validity</li> <li>• Generalizability</li> <li>• Decision Making</li> </ul> <p><i>Screeners only:</i></p> <ul style="list-style-type: none"> <li>• Diagnostic Accuracy</li> </ul> <p><i>PM Tools only:</i></p> <ul style="list-style-type: none"> <li>• Reliability of slope</li> <li>• Validity of slope</li> <li>• Instructional Decision Rules</li> <li>• Improvement Rate Specified</li> <li>• Improvement Rate: End of Year Benchmarks</li> </ul>

# Review Approach

---

## Content

Each assessment was reviewed for content validity. Concepts related to content fell within five domains: (1) early literacy – reading, (2) early literacy – writing, (3) language and communication, (4) health and wellness, and (5) mathematics. Each concept in each domain was scored on two features: (a) whether or not a unique score was provided for that specific concept, on a scale of 0-1, and (b) the extent to which the items on the assessment addressed the concept, on a scale of 0-2.

Weights of each concept were determined by the panel according to the relative perceived importance of that concept within its domain. The maximum score for each concept was 3 points. Once scores were provided, a specified multiplier for each score was applied and the weighted scores were summed to an overall score for each domain. For example, if a concept received a score of 3 and the weight for that score was 1.5, then score for the domain was 4.5 ( $3 \times 1.5=4.5$ ). Finally, domain scores were combined for an overall content coverage score.

The concepts reviewed for each type of assessment are shown in Table 3 along with the weights assigned to each concept.

**Table 3: Scoring Strategy for Content Validity**

		Weight	Pre-K PM	Kindergarten Screening and PM
Emergent Lit: Reading	<b><i>Emergent Literacy - Reading</i></b>			
	Motivation to read	1.0		✓
	PA: syllable segmenting	1.5	✓	✓
	PS: phoneme segmenting and blending	1.5		✓
	PA: initial sounds	1.5	✓	✓
	Alphabet Knowledge: letter names	1.5	✓	✓
	Alphabet Knowledge: letter sounds	1.5		✓
	Concepts of Print: distinguish print elements & direction	1.5	✓	
	Decoding and word recognition	1.5		✓
Comprehension of text read aloud to students	1.25	✓	✓	
Emergent lit: writing	<b><i>Emergent Literacy - Writing</i></b>			
	Motivation to write	1.25		✓
	Writing conventions: first name	1.25	✓	
	Writing conventions: first & last name	1.25		✓
	Writing conventions: letters	1.5	✓	✓
Writing conventions: simple words	1.5		✓	
Lang & communication	<b><i>Language &amp; Communication</i></b>			
	LC: follows single & multistep directions	1.25	✓	✓
	Speech production (intelligible speech)	1.5	✓	✓
	Speaking (conversation skills): verbal & nonverbal	1.5		✓
	Vocab: Expressive vocabulary	1.25	✓	✓
	Vocab: Receptive vocabulary	1.5	✓	✓
	Vocab: Uses common phrases and academic language	1.5	✓	✓
	Speaks in complete sentences (regular complexity)	1.5	✓	✓
Speaks in complete sentences (irregular complexity)	1.5		✓	

**Table 3 (continued): Scoring Strategy for Content Validity**

		Weight	Pre-K PM	Kindergarten Screen & PM
Health and wellness	<b><i>Health and Wellness</i></b>			
	Gross and/or fine motor	1.25	<input type="checkbox"/>	<input type="checkbox"/>
	Self-care	1.25	<input type="checkbox"/>	<input type="checkbox"/>
	Self-awareness/self-regulation	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Relationship skills	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Communicate wishes, feelings, & needs	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Motivation & engagement	1.25	<input type="checkbox"/>	<input type="checkbox"/>
Mathematics	<b><i>Mathematics</i></b>			
	Numeral identification	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Verbal and/or tactile counting	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Adding and/or subtracting	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Geometry and spatial sense language	1.5	<input type="checkbox"/>	<input type="checkbox"/>
	Measurement	1.25		<input type="checkbox"/>
Comparison	1.5	<input type="checkbox"/>	<input type="checkbox"/>	

Because weights were assigned to each concept, and each concept before weighting had a maximum score of 3 points, the maximum domain scores after weighting varies by assessment type, as shown in Table 4.

<b>Table 4: Max Scores for Content Domains After Weighting</b>		
	Pre-K Progress Monitoring	Kindergarten Screening and Progress Monitoring
<b>Emergent Literacy - Reading</b>	22.75	34.75
<b>Emergent Literacy - Writing</b>	9.25	17.5
<b>Language &amp; Communication</b>	26.5	31.0
<b>Health and Wellness</b>	25.75	25.75
<b>Mathematics</b>	23.5	27.25
<b><i>Max Content Score:</i></b>	<b><i>107.5</i></b>	<b><i>136.25</i></b>

## Psychometrics

Following the review of content validity, each assessment was reviewed on its psychometric properties. Psychometric properties under review fell into six broad categories: (1) reliability, (2) validity, (3) generalizability, (4) decision making, (5) diagnostic accuracy, and (6) progress monitoring features.

Each category and the features within each category were scored in the following three ways:

- **Publisher Description:** This score reflects how the publisher describes the evidence for each feature in the proposal.
- **Documentation:** This score reflects the degree to which the publisher provides documentation of the evidence described.
- **Quantitative Evidence:** This score reflects the strength of the quantitative estimate of the psychometric feature based on the evidence provided.

Table 5 shows the maximum scores that could be earned for each category and for each feature within each category. If a cell in Table 5 does not have a max score it means no score was assigned to that category or feature.

Reliability and validity were scored at the category level only on publisher description and documentation. Reviewers considered four types of reliability and three types of validity in assigning scores for reliability and validity. On quantitative evidence for reliability and validity, reviewers assigned scores for each of the four types of reliability and three types of validity, and the total score was the sum of the four reliability scores and the sum of the three validity scores.

Generalizability was scored at the category level for quantitative evidence only. Decision-making was scored at the category level based on the publisher’s description. Diagnostic accuracy was scored at the category level in all three ways: publisher description, documentation, and quantitative evidence.

The category, progress monitoring features, was scored on five features of progress monitoring based on the publisher's description and documentation: (1) reliability of slope, (2) validity of slope, (3) instructional decision rules, (4) specification of improvement rates, and (5) end-of-year benchmarks. Four of these features (all except instructional decision rules) were also scored based on quantitative evidence.

To weight the relative importance of publisher description, documentation, and quantitative evidence, panelists arrived at the following weighting approach: (a) publisher description should reflect 10% of the total score for each category; (b) documentation should reflect 15% of the total score for each category; and (c) quantitative evidence should reflect 75% of the total score for each category. In Table 5, the column labeled Weight Multiplier shows the numerical weight that was used to transform the entered score (0–2 or 0–3) into the final score used in summarizing the assessment.

**Table 5: Scoring Strategy for Psychometrics**

	Publisher Description (10%)		Documentation (15%)		Quantitative Evidence (75%)		<b>Max Score</b>	Included in PM Scoring	Included in Screener Scoring
	Max Score	Weight Multiplier	Max Score	Weight Multiplier	Max Score	Weight Multiplier			
<b>Reliability</b>	<b>3</b>	<b>0.6</b>	<b>3</b>	<b>0.9</b>	<b>12*</b>	<b>1.125</b>	<b>18</b>	✓	✓
Test-retest	-	-	-	-	3	-	NA		
Inter-rater	-	-	-	-	3	-	NA		
Alternate form	-	-	-	-	3	-	NA		
Coefficient alpha	-	-	-	-	3	-	NA		
<b>Validity</b>	<b>2</b>	<b>0.65</b>	<b>2</b>	<b>0.975</b>	<b>9*</b>	<b>1.083</b>	<b>13</b>	✓	✓
Criterion-related validity	-	-	-	-	3	-	NA		
Predictive validity	-	-	-	-	3	-	NA		
Discriminant validity	-	-	-	-	3	-	NA		
<b>Generalizability</b>	-	-	-	-	<b>3</b>	-	<b>3</b>	✓	✓
<b>Decision Making</b>	<b>3</b>	-	-	-	-	-	<b>3</b>	✓	✓
<b>Diagnostic Accuracy (classification)</b>	<b>3</b>	<b>0.267</b>	<b>3</b>	<b>0.4</b>	<b>2</b>	<b>3</b>	<b>8</b>		✓
<b>Progress Monitoring Features</b>	-	-	-	-	-	-	<b>33</b>	✓	
Reliability of slope	2	0.3	2	0.45	2	2.25	6		
Validity of slope	2	0.3	2	0.45	2	2.25	6		
Instructional decision rules**	3	0.8	3	1.2	-	-	6		
Improvement rate specified	2	0.3	2	0.45	2	2.25	6		
End-of-year benchmarks	3	0.3	3	0.45	3	2.25	9		

\*sum of individual category quantitative evidence scores

\*\*quantitative evidence not included; weights are adjusted to 40% for PD & 60% for D

Considering the weights assigned to each score, the maximum psychometric score for each category is shown in Table 6.

<b>Table 6: Max Scores for Psychometrics</b>			
	Pre-K Progress Monitoring	Kindergarten Screening	Kindergarten Progress Monitoring
<b>Reliability</b>	18	18	18
<b>Validity</b>	13	13	13
<b>Generalizability</b>	3	3	3
<b>Decision Making</b>	3	3	3
<b>Diagnostic Accuracy</b>	NA	8	NA
<b>Progress Monitoring Features</b>	33	NA	33
<b><i>Max Psychometric Score:</i></b>	<b>70</b>	<b>45</b>	<b>70</b>

## **Administrative Feasibility**

For assessments on the recommended lists, panelists and CORE project staff rated the assessments together in Meeting 3 for feasibility related to administration. This rating was not included in the final score that determined whether an assessment was recommended or not. The purpose of the ratings is to provide important contextual information schools and districts can use in choosing assessments that fit their local contexts. Six categories of feasibility were identified by CORE and the panel and each category was assigned up to three points (See Table 7 and Appendix G). A percentage score was calculated by dividing the assessment’s administrative feasibility score by 18, the total score possible. The distribution of recommended assessments, based on the administration feasibility percentage score, was then divided into high, medium and low administrative feasibility groups. Additional comments and notes about feasibility of use are also provided in the summary sheets for each assessment.

<b>Table 7: Scoring Strategy for Administrative Feasibility</b>	
<b>Category</b>	<b>Range</b>
Teacher Friendly	0-3
Administrator Friendly	0-3
Student Friendly	0-3
Administration Format	0-3
Language	0-3
Cultural Relevance	0-3
<b>Max Score</b>	<b>18</b>

## **Placement on Lists**

The summed scores for content and for psychometric features were weighted to reflect that 35% of the total score would be based on content and 65% of the total score would be based on psychometrics.

Each assessment's weighted score was transformed into a percentage score based on the total points possible. This overall percentage was used to order the assessments from highest score to lowest score. The highest percentage score achieved was 64.3% of all possible points<sup>2</sup> and the lowest percentage score was 3.5% of all possible points.

CORE used cut-off score that was roughly 0.5 standard deviations below the mean average score of the total number of assessments. This translated to assessments that received at least 25% of the total points possible for content and psychometrics.

Last, only assessments that provided an overall (e.g., composite) domain score for the domains they measure were included.

For kindergarten assessments, this resulted in 14 assessments recommended for use. For prekindergarten assessments, this resulted in 9 assessments recommended for use.

The recommended list of prekindergarten assessments for student progress monitoring is provided in Appendix H. The recommended list of kindergarten student screening and progress monitoring assessments is provided in Appendix I.

Appendices H and I offer a brief interpretation guide as well as additional information about each of the recommended assessments.

**Summary sheets.** Each recommended assessment has an accompanying one-page summary that provides specific information about content coverage and psychometric strengths and limitations, as well as summary information about feasibility of using each assessment and general notes from the panel about each assessment. All recommended assessments received at least 25% of total points possible and provide a score for the content domains that it covers. The way each assessment reached a minimum of 25% of the total possible score may have varied substantially based on different combinations of content and psychometric evidence. Thus, administrative feasibility factors may provide meaningful context about how the different assessments could be used best in different settings. It will be important for schools or districts to choose assessments that represent the best fit for their own local context, and to reviewing the score sheets carefully may be helpful in that process.

---

<sup>2</sup> This indicates that there were no available assessments fully met all of the content and psychometric features that the panel identified for review. It would be highly unusual for an assessment to be able to accomplish each and every of those criteria but it also underscores the need for further development of excellent assessments for this age group.

# Recommendations for Use

---

*The lists provided in the appendices reflect the recommended assessments. Below, CORE provides additional context for overall use and some specific guidance on potential best use of key assessments.*

## General Recommendations

Overall, assessments should be used as specified in administration manuals. Deviations from administration instructions including frequency of assessing, omitting domains, etc., will result in scores that may not reflect a student’s true performance or needs. Assessments marked as recommended for kindergarten screeners should be used for a screening purpose, most commonly at the beginning of the kindergarten school year. Assessments marked as prekindergarten progress monitoring assessments should be used for a progress monitoring purpose, at the intervals recommended by the publisher. When a recommended prekindergarten progress monitoring assessment and a kindergarten screening assessment are “paired” and fit together conceptually and psychometrically, they can be used effectively together as part of a multi-year assessment system. Recommended Spanish and English versions of either prekindergarten progress monitoring or kindergarten screening assessments should be “paired” and used together, based on which language the students are assessed in. Some exceptions or nuances related to individual assessments are noted below.

## Combining Recommended Assessments to Measure Multiple Dimensions

Several high scoring assessments on CORE’s recommended lists were unidimensional, reflecting the fact that unidimensional assessments tend to have more depth but do not have breadth of coverage across domains. This is sometimes reflected unidimensional assessments having relatively strong psychometric properties. However, development through prekindergarten and kindergarten entry represents multiple important dimensions. Therefore, assessments that are multi-dimensional in nature offer advantages that unidimensional assessments lack. Districts and schools should review the content areas (domains) covered by each assessment and consider ways of selecting and combining assessments to make sure they are assessing critical domains in prekindergarten and kindergarten.

## Some Key Principles Regarding the Appropriate Use (and Avoiding the Misuse) of the Recommended Assessments

The recommended assessments are tools for making decisions about instruction and about meeting student needs. The quality of these decisions depends on the appropriate use—and avoiding the unintentional misuse—of these assessments. We address principles of appropriate assessment that fall into three categories: (a) using assessments according to their intended purpose; (b) administering assessments according to correct administration procedures; (c) interpreting assessment results appropriately.

## Intended Purpose

First, all assessments are developed for one or more specific purposes and the recommended pre-k and kindergarten assessments should be used for these purposes only.

The purpose of the recommended pre-k assessments is to monitor the progress of students in different domains during the pre-k year to help determine if the progress they are making is adequately preparing them to be ready for kindergarten. This means that pre-k assessments are intended to be administered at least two times each year and it is the progress, or growth, students make from one assessment to the next, or across several assessments, that is the primary consideration in interpretation of the scores.

The purpose of kindergarten assessments is to screen students for kindergarten readiness. Kindergarten readiness assessments are usually administered at the beginning of the kindergarten year across multiple domains and the results are used to determine the extent to which students are “ready” for kindergarten. This may include pointing to areas that a teacher will need to pay more attention to throughout the year, and in some cases may point to needs for additional assessments or referrals for additional services.

We have noted that several of the kindergarten readiness assessments are also appropriate to administer for progress monitoring purposes<sup>3</sup>. In other words, the assessments can be administered two or more times during the kindergarten year to determine the degree of progress students are making.

## Administering Tests Correctly

In order to use assessments for their intended purpose in pre-k or kindergarten, the assessments must be administered as intended. This means that “standardized” administration procedures are followed for those assessments that are administered directly to students. It also means that assessment data collected in other ways—through teacher observations of students or through the collection of student work samples, for example—are collected in a manner specified in the assessment manual. In some cases, the recommended assessments do not specify all of the details in how the assessment data are to be collected. In these cases, it is important that all the individuals who are collecting the assessment data—e.g., teachers—should know and use the same clearly defined data collection procedures. The highest quality assessments will explain the important details involved in how the data are to be collected and it will be up to the school, district, or assessment team to make sure all individuals who collect data are adequately trained.

## Interpreting Assessment Results

When assessments are used for their intended purposes and are correctly administered, then the results can be interpreted appropriately. The key principle here is making a direct connection between the purpose of the assessment and the interpretation of the results. Consequently, the interpretation of the pre-k results should focus on the progress students are making over the course of the pre-k year in preparing them for kindergarten. Interpreting the results of progress monitoring assessments should answer such questions as (a) Is the student making progress?; (b) Is the progress adequately preparing them for kindergarten?; (c) Does the student’s rate of progress indicate that additional or different instruction or support should be provided to increase progress?

---

<sup>3</sup> See Appendices H through I, the single-page summary of each assessment to see which kindergarten assessments are also appropriate to use as a progress monitor throughout the kindergarten year.

The interpretation of the kindergarten readiness assessments should focus on key screening/readiness questions such as: (a) Is the child demonstrating proficiencies above or below what is expected in kindergarten? (b) Should the student receive additional or different instruction or support to improve their learning and performance in key domain or concepts related to the readiness assessment results? (c) Does the student’s performance on the screening/readiness assessment provide useful information for determining the type of additional instruction or support that would be beneficial?

As mentioned above, some of the kindergarten screening assessments can also be administered more than once to help determine the progress a student is making over time. For kindergarten screening assessments that are also being used as to monitor progress, the same types of interpretation questions raised in the pre-k section above are relevant.

### [Avoiding the Misuse of Assessment Results](#)

Screening/readiness and progress monitoring assessments are primarily intended to identify those students who are scoring below expectations on screening assessments, or not making sufficient progress based on progress monitoring assessments, for the purpose of providing additional instruction and support for those students. Without extensive other data, diagnosing learning disabilities or developmental delays, or making any other “high stakes” diagnostic decisions about students is not appropriate. Similarly, screening and progress monitoring assessments are not designed to evaluate teachers and should not be used for any type of important decision regarding teaching or teacher quality<sup>4</sup>. Last, some kindergarten screening assessments have a decent to strong ability to predict later achievement (for example, standardized scores in elementary), but others do not have this connection well documented. If schools or districts are seeking to predict how current cohorts of kindergarten students will perform in the coming years (for example, to set benchmarks or districtwide goals for performance), the available research about each kindergarten assessment’s correlation and/or *predictive validity* should be carefully considered. Basing benchmark goals on kindergarten scores that have a poor or even moderate ability to predict later achievement will result in false negatives (it will look like students missed the mark when they really hit it) or false positives (it will look like students surpassed the mark when in fact, they did not).

The purpose of progress monitoring and kindergarten screening assessments is to make reasonable, low stakes decisions that involve helping students improve their learning and behavior outcomes on important domain and concepts.

Screening and progress monitoring data can be used in conjunction with other high-quality data to help determine how well education systems are working to support students and teachers. When there are systems-level difficulties—for example, not enough time is being spent on teaching literacy—screening and progress monitoring data aggregated in reasonable ways can to help provide information about these types of potential systems-level problems. However, these data should not be used on their own to evaluate programs or teachers, and they should not be used to make high stakes decisions involving students and families.

---

<sup>4</sup> For additional resources on this topic, see: <https://www.theounce.org/wp-content/uploads/2017/03/PolicyConversationKRA2017.pdf>

# References

---

**CORE conducts an ongoing review of emergent literature on issues relating to early childhood. We drew on multiple resources to help inform the development of our scoring rubric and overall review process.**

- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25–50.
- Barnett, W. S., Friedman-Krauss, A. H., Gomez, R., Horowitz, M., Weisenfeld, G. G., Brown, K. C., & Squires, J. H. (2016). *The state of preschool 2015: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Bassok, D., Latham, S. (2016). *Kids Today: Changes in school-readiness in an early childhood era*. Ed Policy Works, Charlottesville, VA.
- Bierman, K. L., Domitrovich, C. E., Nix, R., Gest, S., Welsh, J. A., Greenberg, M. T., Blair, C., et al. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79*, 1802–1817.
- Bredenkamp, S., & Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs*. Washington, DC: National Association for the Education of Young Children.
- Burchinal, M., Vernon-Feagans, L., Vitiello, V., Greenberg, M., & Family Life Project Key Investigators. (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early Childhood Research Quarterly, 29*(1), 41–51.
- Burchinal, M., Kainz, K. & Cai, Y. (2011). How well are our measures of quality predicting to child outcomes: A meta-analysis and coordinated analysis of data from large scale studies of early childhood settings. In Zaslow, M., Tout, K., Halle, T., & Martinez-Beck I. (Eds.), *Next steps in the measurement of quality in early childhood settings*. Baltimore: Brookes Publishing.
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166–176.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579–620.
- Campbell, A.B., Denham, S.A., Howarth, G.Z., Jones, S.M., Whittaker, J.V., Williford, A.P., Wiloughby, M.A., Yudron, M., & Draling-Churchill, K. (2016). Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique and recommendations. *Journal of Applied Developmental Psychology, 45*, 19-41.
- Chernoff, J. J., Flanagan, K. D., McPhee, C., & Park, J. (2007). *Preschool: First findings from the Preschool Follow-Up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). First Look*. NCES 2008-025. *National Center for Education Statistics*.
- Chien, N. C., Howes, C. Pianta, R., Burchinal, M., Ritchie, S., Bryant, D.M., Clifford, R.M., Early, D.M., & Barbarin, O.A. (2010). Children’s classroom engagement and school readiness gains in prekindergarten. *Child Development, 81*, 5, 1534-1549.
- Golan, S., Woodbridge, M., Davies-Mercier, B., Pistorino, C. (2016). *Case studies of the early implementation of kindergarten entry assessments*. Policy and Program Studies Service, U.S. Department of Education, Washington, DC.
- Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short and long-term benefits of educational opportunity. *School Psychology Quarterly, 16*(1), 9–30.
- Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology, 41*(6), 872–884.
- Halle, T.G. & Darling-Churchill, K.E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology, 45*, 8-18.

- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76(5), 949–967.
- Heckman, J., Pinto, R., & Savelyev, P. A. (2012). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes (NBER Working Paper No. 18581). Cambridge, MA: National Bureau of Economic Research. Retrieved from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3951747/>.
- Kame'enui, E.J., Fuchs, L., Francis, D.J., Good, R., O'Connor, D., Simmons, G.T., & Torgesen, J.K. (2006). The adequacy of tools for assessing reading competence: A framework and review. *Educational Researcher*, 35, 4, 3-11.
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, 104(5), 409–426.
- Morrison, F. J., & Connor, C. M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology*, 40(6), 493–500.
- Neuman S. B., Wright T. S. (2010). Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching. *The Elementary School Journal*, 111, 63–86.
- NICHD Early Child Care Research Network (2005). Early child care and children's development in the primary grades: Follow-up results from the NICHD study of early child care. *American Educational Research Journal*, 42(3), 537–570.
- Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, C., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9(3), 144–159.
- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... & Ciarico, J. (2010). Head Start Impact Study. Final Report. *Administration for Children & Families*.
- Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state prekindergarten programs. *Journal of Policy Analysis and Management*, 27, 122–154.
- Weisenfeld, G.G. (2017). Assessment tools used in kindergarten entry assessments (KEAs) State Scan. Center on Enhancing Early Learning Outcomes, Fast Fact, January 2017.
- Reardon, S.F. & Portilla, X.A. (2016). Recent trends in oncome, racial and ethnic school readiness gaps at kindergarten entry. American Educational Research Association Open.
- Regenstein, E., Connors, M., Romero-Juardo, R., & Weiner, J. (2017). Uses and misuses of kindergarten readiness assessment results. *The Ounce of Prevention; Policy Conversations*, 6, 1, 1-48.

# Appendices

---

A: List of Panelists

B: Standards Alignment

C: State of Assessments – Pre-K and Kindergarten

D: Requests for Information (RFIs) – Pre-K and Kindergarten

E: Rubric and Glossary

F: Master Scoring Sheet

G: Tool for Reviewing Administrative Feasibility

H: Final List of Recommended Pre-K Assessments

I: Final List of Recommended Kindergarten Assessments



**SMU** | ANNETTE CALDWELL SIMMONS  
SCHOOL OF EDUCATION & HUMAN DEVELOPMENT

CENTER ON RESEARCH & EVALUATION

