

Early Childhood Assessment Review: Glossary & Rubric

Instructions: Assessments will first be reviewed for content validity. The results of content validity will dictate the degree to which assessments are reviewed for psychometric features. Use the “Review Matrix” tool to assign scores. Use a new copy of the tool per assessment.

Initial Questions:

1. Is the tool being reviewed for the *pre-k* progress monitoring list, for the *kindergarten* list, or *both*?
2. If it is being reviewed for kindergarten, is it a *screener* only, a *progress monitor* only, or *both*?

Screeners: 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.

Progress Monitor: 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 calculated as a slope of improvement score based on two or more administrations of the measure. The slope estimate is used as the basis for determining the adequacy of student progress, typically in relation to external progress criteria.

Administrative	Content	Psychometric
<ul style="list-style-type: none"> Title Publisher Recommended Use Price Per Student Format: Direct or Observation Format: Group or 1-on-1 Scoring Language Grade Levels Test Format (e.g., pp) Requirements to Admin Time Requirements Per Student 	<p><i>Content Validity Across 5 Domains:</i></p> <ol style="list-style-type: none"> 1. Emergent Literacy – Reading (5 concepts) 2. Emergent Literacy- Writing (3 concepts) 3. Language & Communication (6 concepts) 4. Health & Wellness (6 concepts) 5. Mathematics (5 concepts) 	<p><i>All:</i></p> <ul style="list-style-type: none"> Reliability Validity Generalizability Decision-Making <p><i>Screeners Only:</i></p> <ul style="list-style-type: none"> Diagnostic Accuracy <p><i>Progress Monitoring (PM) Only:</i></p> <ul style="list-style-type: none"> Reliability of Slope Validity of Slope Instructional Decision Rules Improvement Rate Specified Improvement Rate: End of Year Benchmarks



Administrative Features

CORE will complete and review this information with administrators and practitioners on the panel. No scoring required.

Construct	Feature	Definition
Administrative Features	Title	Title of assessment.
	Publisher	Name of publisher.
	Recommended Use	The stated recommended use of the assessment tool (e.g., instructional, research, and standardized applicability).
	Price Per Student	The cost of administering the tool to one child per academic year.
	Administrative Format	Role of the student(s) and test administrator (group format, direct one-on-one, observation).
	Scoring	Scores available: raw score, scale score, normed scores, category of performance, etc.
	Language	Languages available (English, Spanish, alternate form Spanish).
	Grade Levels	Is the test aligned across multiple grade levels?
	Test Format	Physical format: paper pencil, online, computer adaptive, etc.
	Requirements to Admin	Are any specialized certifications required to administer the assessment?
	Time Requirements Per Student	Amount of time needed to administer the assessment once to one student.
	Score Report Formats	The type of score reports made available (e.g., parent, teacher, school-wide, classroom).



Content

Panelists with content and assessment expertise will review assessments for content validity.

Scoring

Part 1: Is a score given? (applies to domains & concepts)	Part 2: Depth of coverage? (applies to concepts only)
<ul style="list-style-type: none"> • Yes (1) = The assessment generates a unique score for that specific content area or domain. • No (0) = The assessment does not generate a unique score for that specific content area or domain. <p><i>*A "score" for a concept may be a sub-scale score representing multiple items, or it may be a single item, depending on the concept & panel agreement.</i></p>	<ul style="list-style-type: none"> • 2 = The assessment overall and individual items strongly represent key aspects of the concept. • 1 = The assessment overall and individual items moderately represent key aspects of the concept. • 0 = The assessment overall and individual items do not at all represent key aspects of the concept.

Domain	Concept	Scoring	
		Pre-K	Kinder
Emergent Literacy - Reading	Motivation to read.		•
	PA: syllable segmenting.	•	•
	PS: phoneme segmenting and blending.		•
	PA: initial sounds.	•	•
	Alphabet Knowledge: letter names.	•	•
	Alphabet Knowledge: letter sounds.		•
	Concepts of Print: distinguish print elements & direction.	•	
	Decoding and word recognition.		•
	Comprehension of text read aloud to students.	•	•
Emergent Literacy - Writing	Motivation to write.		•
	Writing conventions: first name.	•	
	Writing conventions: first & last name.		•
	Writing conventions: letters.	•	•
	Writing conventions: simple words.		•
Language and Communication	LC: follows single & multistep directions	•	•
	Speech production (intelligible speech)	•	•
	Speaking (conversation skills): verbal & nonverbal		•
	Vocab: Expressive vocabulary	•	•
	Vocab: Receptive vocabulary	•	•
	Vocab: Uses common phrases and academic language	•	•
	Speaks in complete sentences (regular complexity)	•	•
	Speaks in complete sentences (irregular complexity)		•
Health and Wellness	Gross and/or fine motor	•	•
	Self-care	•	•
	Self-awareness/self-regulation.	•	•
	Relationship Skills.	•	•
	Communicate wishes, feelings, & needs.	•	•
	Motivation & Engagement.	•	•
Mathematics	Numeral Identification	•	•
	Verbal and/or tactile counting.	•	•
	Adding and/or subtracting.	•	•
	Geometry and spatial sense language.	•	•
	Measurement		•
	Comparison	•	•



Psychometric Features

Publisher Description: This score is reflective of how the publisher describes the evidence for each feature in the proposal.

Documentation: This score is reflective of the degree to which the publisher provides reasonable documentation of the evidence described.

Quantitative Evidence: This is the strength of the actual quantitative evidence provided by the documentation and/or description.

*Score at the domain or concept level, whichever is most appropriate and matched to the level at which the publisher provides psychometric information (e.g., if features are only described at the "whole assessment" level, score psychometrics for the whole assessment).

Constructs Applicable to ALL Measures

Construct	Feature	Definition	Scoring		
			Publisher Description	Documentation	Quantitative Evidence
Reliability	Test Reliability	Reliability refers to the consistency of the student's test score or how the student is classified from one test administration or rating to the next. A test is considered reliable if it produces the same result when administered under different conditions, at different times, or using different forms. A test is also considered reliable if there is internal consistency among the items on the test.	3 = strong evidence: reliability evidence is described for two or more areas; including overall score (e.g., test-retest) and test items (e.g., coefficient alpha); 2 = moderate evidence: reliability evidence is described for two or more areas but reliability evidence lacking in either overall score (e.g., test-retest) or among test items (e.g., coefficient alpha); 1 = minimal evidence: reliability evidence is described for one area; 0 = no evidence: reliability evidence is not described.	3 = strong evidence: reliability evidence is described for two or more areas; including overall score (e.g., test-retest) and test items (e.g., coefficient alpha); 2 = moderate evidence: reliability evidence is described for two or more areas but reliability evidence lacking in either overall score (e.g., test-retest) or among test items (e.g., coefficient alpha); 1 = minimal evidence: reliability evidence is described for one area; 0 = no evidence: reliability evidence is not described.	Sum of Quantitative Evidence scores for each feature within the <u>validity</u> construct.
	Test-Retest Reliability	Test administered at different points in time.	N/A	N/A	3 = Strong: reliability estimates are greater than .80 2 = Moderate: reliability estimates are between .70 and .79 1 = Minimal: reliability estimates are below .70 0 = None: reliability estimates are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.
	Inter-Rater Reliability	Score consistency among test administrators.	N/A	N/A	3 = Strong: reliability estimates are greater than .80 2 = Moderate: reliability estimates are between .70 and .79 1 = Minimal: reliability estimates are below .70 0 = None: reliability estimates are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.
	Alternate Form	Different forms or versions of the same test. (whichever is applicable: alternate forms for screeners (e.g., form a and b given at the same time to different kids; for progress monitoring (e.g., multiple time points))	N/A	N/A	3 = Strong: reliability estimates are greater than .80 2 = Moderate: reliability estimates are between .70 and .79 1 = Minimal: reliability estimates are below .70 0 = None: reliability estimates are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.
	Internal Consistency	The internal consistency of the test. This may include coefficient alpha, standard error, or properties of item response theory (computer adaptive assessments).	N/A	N/A	3 = Strong: reliability estimates are greater than .80 2 = Moderate: reliability estimates are between .70 and .79 1 = Minimal: reliability estimates are below .70 0 = None: reliability estimates are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.

Construct (cont.)	Feature (cont.)	Definition (cont.)	Publisher Description (cont.)	Documentation (cont.)	Quantitative Evidence (cont.)
Validity	Test Validity	The extent to which an assessment accurately measures what it is supposed to measure.	2 = strong evidence: validity evidence is described for two or more areas of validity. 1 = moderate evidence: validity evidence is described for one area.	2 = strong evidence: validity evidence is described for two or more areas of validity. 1 = moderate evidence: validity evidence is described for one area.	Sum of Quantitative Evidence scores for each feature within the <u>reliability</u> construct.
	Criterion-Related Validity	The extent to which the assessment score or classification is related to relevant outcomes assessed at approximately the same time.	N/A	N/A	3 = Strong: correlations with other relevant outcome measures are typically above .70 2 = Moderate: correlations with other relevant outcome measures are typically between .50 and .70 1 = Minimal: correlations with other relevant outcome measures are inconsistent and include correlations below .50 0 = None: correlations are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.
	Predictive Validity	The extent to which the assessment score or classification is related to relevant outcomes assessed in the future.	N/A	N/A	3 = Strong: correlations with other relevant outcome measures are typically above .60 – .70 2 = Moderate: correlations with other relevant outcome measures are typically between .40 and .60 1 = Minimal: correlations with other relevant outcome measures are inconsistent and include correlations below .40 0 = None: correlations are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.
	Discriminant Validity	The extent to which the assessment score or classification is not related to constructs not being assessed.	N/A	N/A	3 = Strong: correlations are consistently and clearly higher for assessments related to the target construct than on assessments not related to the target construct. 2 = Moderate: correlations are inconsistently and only somewhat higher on assessments related to the target construct than on assessments not related to the target construct. 1 = Minimal: correlations are similar on assessments related to the target construct than on assessments not related to the target construct. 0 = None: correlations on constructs not related to the target construct are not provided. N/A = Not Applicable: estimates are not applicable to this assessment.
Generalizability	N/A	The sample of students on which the measure was studied is similar to Texas students. A measure is more generalizable if studies have been conducted on larger, more representative samples.	N/A	N/A	3 = Strong: large representative sample. 2 = Moderate: moderately sized representative sample. 1 = Minimal: limited representative sample. 0 = No evidence: sample size is not provided. N/A = Not Applicable: estimates are not applicable to this assessment.

Construct (cont.)	Feature (cont.)	Definition (cont.)	Publisher Description (cont.)	Documentation (cont.)	Quantitative Evidence (cont.)
Decision Making	N/A	Evidence that the use of the results of the measure leads to improvements in education decision making, including instructional planning, instruction provided, or student outcomes.	<p>3 = strong basis for link is described between score(s) and improved instructional planning, instruction, or student outcomes.</p> <p>2 = moderate basis for link is described between score(s) and improved instructional planning, instruction, or student outcomes.</p> <p>1 = weak basis for link is described between score(s) and improved instructional planning, instruction, or student outcomes.</p> <p>0 = no basis for link is described between score(s) and improved instructional planning, instruction, or student outcomes.</p>	N/A	N/A

Constructs Applicable to Screening Measures ONLY

Construct	Feature	Definition	Publisher Description	Scoring	
				Documentation	Quantitative Evidence
Diagnostic Accuracy	Classification Accuracy	<p>The extent to which a screening tool is able to accurately classify students into “at risk” and “not at risk” categories.</p> <ul style="list-style-type: none"> Ô An appropriate outcome measure was used. Ô Students in the study were an appropriate sample for determining risk (i.e., in relation to Texas students; and receiving regular or general instruction vs. specialized instruction). Ô Risk was adequately defined (e.g., below the 20th %-tile). Ô Classification analyses with cut-points were adequately conducted. 	<p>3 = strong: all considerations are described.</p> <p>2 = moderate: most (3) considerations are described.</p> <p>1 = limited: 1 or 2 considerations are described.</p> <p>0 = none: no considerations are described.</p>	<p>3 = strong: evidence is provided for all considerations.</p> <p>2 = moderate: evidence is provided for most (3) considerations.</p> <p>1 = limited: evidence is provided for 1 or 2 considerations.</p> <p>0 = none: evidence is not provided for any considerations.</p>	<p>Area Under the Curve (AUC) Statistic: an overall indication of the diagnostic accuracy of a Receiver Operating Characteristic (ROC) curve. ROC curves are a generalization of the set of potential combinations of sensitivity and specificity possible for predictors. AUC values closer to 1 indicate the screening measure reliably distinguishes among students who are at risk and who are not at risk. Values at .50 indicate the predictor is no better than chance.</p> <p>2 = strong: AUC is above .85</p> <p>1 = limited: AUC is between .70 and .84</p> <p>0 = none: AUC is below .70 OR is not provided at all.</p> <p>N/A = Not Applicable: estimates are not applicable to this assessment.</p>

Constructs Applicable to Progress Monitoring Measures ONLY

Construct	Feature	Definition	Scoring		
			Publisher Description	Documentation	Quantitative Evidence
Reliability	Reliability of Slope	Extent to which the slope of improvement accurately represents rate of improvement.	<p>2 = strong evidence: reliability of slope evidence and quantitative evidence is described.</p> <p>1 = moderate evidence: reliability of slope evidence is described but quantitative evidence is not described.</p> <p>0 = no evidence: reliability of slope evidence is not described.</p>	<p>2 = strong evidence: reliability of slope evidence and quantitative evidence is provided.</p> <p>1 = moderate evidence: reliability of slope evidence is provided but quantitative evidence is not provided.</p> <p>0 = no evidence: reliability of slope evidence is not provided.</p>	<p>2 = strong: reliability of slope estimates are consistently greater than .70</p> <p>1 = moderate: reliability of slope estimates are consistently less than .70</p> <p>0 = none: reliability of slope estimates are not provided.</p> <p>N/A = Not Applicable: estimates are not applicable to this assessment.</p>
Validity	Validity of Slope	The extent to which the slope of improvement corresponds to level of performance on important outcomes.	<p>2 = strong evidence: validity of slope of improvement evidence is described and quantitative estimates of correspondence with level of performance outcomes are described.</p> <p>1 = moderate evidence: validity of slope of improvement evidence is described but quantitative estimates of correspondence with level of performance outcomes are not described. 0 = no evidence: validity of slope of improvement</p>	<p>2 = strong evidence: validity of slope of improvement evidence is provided and quantitative estimates of correspondence with level of performance outcomes are provided.</p> <p>1 = moderate evidence: validity of slope of improvement evidence is provided but quantitative estimates of correspondence with level of performance outcomes are not provided. 0 = no evidence: validity of slope of improvement evidence</p>	<p>2 = strong: slope of improvement estimates are correlated with level of performance outcomes at .60 or higher.</p> <p>1 = moderate: slope of improvement estimates are correlated with level of performance outcomes are below .60</p> <p>0 = none: slope of improvement correlations with level of performance outcomes are not provided.</p> <p>N/A = Not Applicable: estimates are not applicable to this assessment.</p>
Instructional Decision Rules	N/A	The measure includes clear decision rules linking student performance to needed changes to instruction to increase performance or the need to increase the student's goals on the measure.	<p>3 = strong basis for decision rule link is described between student performance and the need to change instruction or increase the student's goals. 2 = moderate basis for decision rule link is described between student performance and the need to change instruction or increase the student's goals. 1 = weak basis for decision rule link is described between student performance and the need to change instruction or increase the student's goals. 0 = no basis for decision rule link is described.</p>	<p>3 = strong evidence is presented demonstrating that use of decision rules to change instruction or increase student goals leads to improved student outcomes.</p> <p>2 = moderate evidence is presented demonstrating that use of decision rules to change instruction or increase student goals leads to improved student outcomes.</p> <p>1 = weak evidence is presented demonstrating that use of decision rules to change instruction or increase.</p> <p>0 = no evidence is presented demonstrating that use of decision rules to change instruction or increase student goals leads to improved student outcomes.</p>	N/A

Construct (cont.)	Feature (cont.)	Definition (cont.)	Publisher Description (cont.)	Documentation (cont.)	Quantitative Evidence (cont.)
Improvement Rates	Improvement Rate Specified	Rates of improvement specify the slopes of improvement or average increases, based on a line of best fit through the student's scores.	<p>2 = strong evidence: rates of improvement is described and quantitative data specifying rates of improvement is described.</p> <p>1 = moderate evidence: rates of improvement is described but quantitative data specifying rates of improvement are not described.</p> <p>0 = no evidence:</p>	<p>2 = strong evidence: rates of improvement is presented and quantitative data specifying rates of improvement is presented.</p> <p>1 = moderate evidence: rates of improvement are presented but quantitative data specifying rates of improvement are not presented.</p> <p>0 = no evidence:</p>	<p>2 = evidence is clear that slope of improvement or average increase is calculated.</p> <p>1 = evidence is somewhat clear that slope of improvement or average increase is calculated.</p> <p>0 = evidence is not clear that slope of improvement or average increase is calculated.</p> <p>N/A = Not Applicable: estimates are not applicable to this assessment.</p>
	End-of-Year Benchmarks	End-of-year benchmarks specify the level of performance expected by grade (by pre-k year or in K) at the end of the academic year.	<p>3 = strong evidence: end of year benchmark is discussed, benchmark level of performance at end of year is specified, and validity evidence for benchmark level of performance at end of year is described.</p> <p>2 = moderate evidence: end of year benchmark is discussed, benchmark level of performance at end of year is specified, but validity evidence for benchmark level of performance at end of year is not described.</p> <p>1 = minimal evidence: end of year benchmark is discussed, but benchmark level of performance at end of year is not specified, and validity evidence for benchmark level of performance at end of year is not described.</p> <p>0 = no evidence:</p>	<p>3 = strong evidence: end of year benchmark is presented, benchmark level of performance at end of year is specified, and validity evidence for benchmark level of performance at end of year is provided.</p> <p>2 = moderate evidence: end of year benchmark is provided, benchmark level of performance at end of year is specified, but validity evidence for benchmark level of performance at end of year is not provided.</p> <p>1 = minimal evidence: end of year benchmark is provided, but benchmark level of performance at end of year is not specified, and validity evidence for benchmark level of performance at end of year is not provided.</p> <p>0 = no evidence: end of year benchmark performance is not provided.</p>	<p>3 = Strong: end of year benchmark score is correlated with other relevant outcome measures above .70</p> <p>2 = Moderate: end of year benchmark score is correlated with other relevant outcome measures between .50 and .70</p> <p>1 = Minimal: end of year benchmark score is correlated with other relevant outcome measures below .50</p> <p>0 = no evidence: correlations with outcome measures are not provided.</p> <p>N/A = Not Applicable: estimates are not applicable to this assessment.</p>