2024-2027 Commissioner's List of Approved Prekindergarten Assessment Instruments

Matrix and Glossary

Administrative Features	Content Features	Psychometric Features
• Title • Publisher • Languages • Age Levels • Time points • Format: Physical • Format: Administration • Training Requirements • Scoring Method • Available Scores • Score Reports • Time per student • Price per student	• Depth of 2022 Prekindergarten Outcome Coverage <u>Scored for Review Calculations:</u> • Gross and Fine Motor Skills • Self-Regulation (behavior, emotion, and attention) • Listening Comprehension • Vocabulary • Phonological Awareness • Alphabet Knowledge • Conventions in Writing • Number Sense • Classifications and Patterns <u>Rated for Presentation Purposes Only:</u> • Personal Health and Safety • Self-Concept • Relationships with Others • Social Awareness • Speaking (Conversation) • Articulation • Sentences and Structure • Comprehension of Text • Concepts of Print • Motivation to Read • Motivation to Write • Writing as a Process • Joining and Separating • Geometry and Spatial Sense • Measurement	• Feasibility • Generalizability • Reliability • Validity • Growth/Improvement

Administrative Features

Recorded (not scored):

- Title (Name of Instrument)
- Publisher (Name of company publishing the instrument)
- Languages
    - English
    - Spanish
    - Language Neutral {publisher indicates that observations can be made on children regardless of the language used by the student, including language switching}
- Age Levels (Required: must address some portion of the 2-5-year-old age range)
- Time points
    - Beginning of Year, Middle of Year, End of Year (BOY, MOY, EOY),
    - Other (Required = 3 times per year)
- Format: Physical
    - Paper/pencil, Computer (not adaptive), Computer adaptive, Survey, Observation
- Format: Administration
    - One-on-one, Group, Observation
- Training Requirements
    - Required - Yes/No; Minimum amount of time required; Required certifications
- Scoring Method
    - Manually (paper), Web-based entry after administration, Computer automated (administered/automatically scored on a computer)
- Available Scores
    - Raw, Percent correct, Scaled, Standard, Percentile, Performance category, Other
- Score Reports
    - Individual, Whole class, Whole school, Parent, Other
- Time per student (Time required to administer the assessment once)
    - Required: $\leq$20 min per domain; $\leq$ 100 min for the entire instrument required
- Price per student (Cost of assessment per student annually)

Recorded (not scored):

Required Criteria	Yes or No
Offered in English and Spanish	
Intended for progress monitoring use in Prekindergarten	
Administered three times a year (Beginning, Middle, and End of Year)	
Age levels appropriate for Prekindergarten	
Administration time is ≤ 20 minutes per student per domain	
Administration time for cumulative test is ≤ 100 minutes per student	
Individually administered to each student	
English and Spanish versions assess the same domains and skills within domains	
Normative/technical data must be no more than 15 years old (i.e., 2008 or later)	
All required skills within each 2022 Texas Prekindergarten domain are assessed	
Health and Wellness	
Emergent Literacy - Language and Communication	
Emergent Literacy - Reading	
Emergent Literacy - Writing	
Mathematics	
Preferred But Not Required	
Parent reports	
Instructional resources for teachers	
Instructional resources for parents	

Content Features

Depth of 2022 Texas Prekindergarten Outcome Coverage

Scoring:

Each required skill from the *2022 Texas Prekindergarten Guidelines: PK3 and PK4 Comprehensive Guide* (henceforth 2022 Texas PK Outcomes) was scored based on the degree to which the instrument addresses the main "gist" of the skills described in the 2022 Texas PK Outcomes. Raters used the 2022 Texas PK Outcomes statements and took into consideration the child behavior examples provided.

Depth of Coverage Rubric

4 Very strongly addresses key aspects of the skill
3 Strongly addresses key aspects of the skill
2 Moderately addresses key aspects of the skill
1 Minimally addresses key aspects of the skill
0 Does not address key aspects of the skill

Skills within each domain required to be considered for inclusion in the 2024-2027 Commissioner's List of Approved Prekindergarten Assessment Instruments are presented in **bold** in the following table. All other skills were also rated, but those scores were not included in the scoring for inclusion in the Commissioner's List. Scores for non-required skills are presented for informational purposes only.

2022 Texas Prekindergarten Domains and Skills	Score	Included in Calculations
Health and Wellness		
Gross and Fine Motor Development	**(0-4)**	**Yes**
Personal Health and Safety	(0-4)	No
Self-Regulation	**(0-4)**	**Yes**
Self-Concept	(0-4)	No
Relationships with Others	(0-4)	No
Social Awareness	(0-4)	No
Emergent Literacy: Language and Communication		
Listening Comprehension	**(0-4)**	**Yes**
Vocabulary	**(0-4)**	**Yes**
Speaking (Conversation)	(0-4)	No
Articulation	(0-4)	No
Sentences and Structure	(0-4)	No
Emergent Literacy: Reading		
Phonological Awareness	**(0-4)**	**Yes**
Alphabet Knowledge	**(0-4)**	**Yes**
Comprehension of Text	(0-4)	No
Concepts of Print	(0-4)	No
Motivation to Read	(0-4)	No
Emergent Literacy: Writing		
Conventions in Writing	**(0-4)**	**Yes**
Motivation to Write	(0-4)	No
Writing as a Process	(0-4)	No
Mathematics		
Number Sense	**(0-4)**	**Yes**
Classification and Patterns	**(0-4)**	**Yes**
Joining and Separating	(0-4)	No
Geometry and Spatial Sense	(0-4)	No
Measurement	(0-4)	No

Psychometric Features

Recorded (not scored):
Reviewers record the years of collection for each psychometric sample and indicate whether the normative and technical data is no more than 15 years old (2008+). Normative and technical data is required to be no more than 15 years old. Instruments where all, or the vast majority, of the data were collected from prekindergarten students prior to 2008, will not be considered for the Commissioner's list.

FEASIBILITY

Scoring: Score using the rubric below.

Feasibility	Components to consider	Rating Description	Rating
Teacher Friendly	• Administration time is manageable. • Administration training requirements are minimal. • Minimal additional materials are required for administration. • Scores and score reports are immediately available. • Scores are easy to interpret. • Score reports for parents are easily generated. • Aggregated score reports are easily generated (e.g., groups, skills, whole class).	3 = Strong 2 = Moderate 1 = Minimal 0 = No evidence	
Student Friendly	• Time requirement is manageable. • Directions and tasks are easy to understand. • Assessment is visually appealing. • Assessment is engaging.	3 = Strong 2 = Moderate 1 = Minimal 0 = No evidence	
Administrator Friendly	• Administration training requirements are minimal. • Scoring requires minimal time. • Scores are easy to interpret. • Score reports for parents are easily generated. • Aggregated score reports are easily generated (e.g., whole class, whole school).	3 = Strong 2 = Moderate 1 = Minimal 0 = No evidence	

GENERALIZABILITY

This is the degree to which the sample(s) of students used to develop the assessment and establish psychometric properties is sufficiently large and demographically similar to the Texas student population.

For each sample employed, score in the following manner:

Sample Size	**Representativeness**
3 = Large	3 = Representative
2 = Moderate	2 = Relatively representative
1 = Limited	1 = Not Very representative
0 = Not provided	0 = Not provided

Scores across samples are combined and averaged. The resulting average score is interpreted on the following scale:

3 = Strong
2 = Moderate
1 = Minimal
0 = No evidence

RELIABILITY

Reliability is the consistency with which scores on a measurement instrument measure an underlying construct. A construct is a trait, an ability, or a behavior that cannot be seen. The trait, ability, or behavior is thought to be responsible for a student's response to a test question.

- Not all aspects of reliability will be applicable to all assessments.
- All instruments should report some type of internal consistency.
- All instruments should report some type of test-test reliability.
- If instrument administrators make some type of determination in order to record a "score" student responses or abilities (e.g., making ratings or indicating correct and incorrect), then some form of inter-rater reliability should be provided.
- If different versions of an assessment are available (e.g., form A, form B, etc.), then some type of alternate form reliability data should be provided.

Internal Consistency: This may include coefficient alpha, standard error, or Item Response Theory (IRT), etc.

Score Value	Evidence
3 = Strong	Majority of estimates are greater than .80
2 = Moderate	Majority of estimates are between .70 and .79
1 = Minimal	Majority of estimates are below .70
0 = No evidence	Estimates are not provided
NA = Not Applicable	Estimates are not applicable to this assessment

Test-Retest Reliability: Test administrations at different points in time.

Score Value	Evidence
3 = Strong	Majority of estimates are greater than .80
2 = Moderate	Majority of estimates are between .70 and .79
1 = Minimal	Majority of estimates are below .70
0 = No evidence	Estimates are not provided
NA = Not Applicable	Estimates are not applicable to this assessment

Inter-rater Reliability: Consistency of scores between different test administrators. Typically, this is measured in reference to multiple administrators assessing the same students.

Score Value	Evidence
3 = Strong	Majority of estimates are greater than .80
2 = Moderate	Majority of estimates are between .70 and .79
1 = Minimal	Majority of estimates are below .70
0 = No evidence	Estimates are not provided
NA = Not Applicable	Estimates are not applicable to this assessment

Alternate Form: a.k.a. - Parallel Forms: Different forms or versions of the same test designed to be equivalent.

Score Value	Evidence
3 = Strong	Majority of estimates are greater than .80
2 = Moderate	Majority of estimates are between .70 and .79
1 = Minimal	Majority of estimates are below .70
0 = No evidence	Estimates are not provided
NA = Not Applicable	Estimates are not applicable to this assessment

VALIDITY

All instruments should provide some type of Construct Validity (Concurrent/Convergent or Discriminative) and Predictive Validity.

Construct Validity: This is the extent to which the score or classification is related to other relevant measures/criteria measured at the same time. Construct validity encompasses concurrent, convergent, and/or discriminative validity, and one or more of these may be reported.

Score Value	Evidence
3 = Strong	estimates with other relevant outcome measures are typically above .70
2 = Moderate	estimates with other relevant outcome measures are typically between .50 and .70
1 = Minimal	estimates with other relevant outcome measures are inconsistent and include correlations below .50
0 = No evidence	estimates are not provided
NA = Not Applicable	estimates are not applicable to this assessment

Predictive Validity: The extent to which the score or classification predicts other relevant measures/criteria measured at a future time.

Score Value	Evidence
3 = Strong	estimates with other relevant outcome measures are typically above .70
2 = Moderate	estimates with other relevant outcome measures are typically between .50 and .70
1 = Minimal	estimates with other relevant outcome measures are inconsistent and include correlations below .50
0 = No evidence	estimates are not provided
NA = Not Applicable	estimates are not applicable to this assessment

Growth/Improvement

The degree to which the instrument is sensitive to growth or improvement.

Score Value	Evidence
3 = Strong	Provides strong evidence of ability to detect growth/improvement over time
2 = Moderate	Provides moderate evidence of ability to detect growth/improvement over time
1 = Minimal	Provides minimal evidence of ability to detect growth/improvement over time
0 = No evidence	Provides no evidence of ability to detect growth/improvement over time