Overview

The Instrument Selection Process includes four phases. The purpose of this optional Instrument Selection Tool is to guide Local Education Agencies (LEAs) through Phase 3 of the Instrument Selection Process. **Phases one, two, and four are expected to be completed outside of the use of this tool.**

Instrument Selection Process

Phase 1	**Awareness and Interest:** Become aware of TEA's data collection requirements and seek information about the instruments on the pre-k Commissioner's List. **Communicate Need:** Define and articulate the need or reason for change of an instrument.
Phase 2	**Define Non-negotiables:** Determine the non-negotiables such as budget, formats, language, required supports, resources, reports, and data sharing. **Identify Committee:** Select committee for the instrument review process.
Phase 3	**Filter the Instruments:** Narrow the number of instruments for the selection committee to review by consulting the identified non-negotiables. Use Step 1 in the optional Instrument Selection Tool or another process. **Evaluate the Instruments:** Selection committee evaluates the instruments using Step 2 in the optional Instrument Selection Tool or another process.
Phase 4	**Trial and Activation:** Selection committee further examines and if possible tries out the product or peruses vendor websites to determine if it will meet their needs and to make a final selection. **Confirm Instrument:** Selection committee confirms the instrument and addresses the next steps necessary for adoption and implementation.

Directions for Phase 3

1. Filter the Instruments: Use the Instrument Analyzing Questions in the Step 1 tab in the Instrument Selection Tool, to narrow the number of instruments for the selection committee to review, by consulting the identified non-negotiables (i.e. budget, formats, language, reports, and data sharing). The Step 1 tab includes the Commissioner's List of Approved Prekindergarten Progress Monitoring Instruments listed in alphabetical order and Instrument Analyzing Questions to determine which instruments will be further evaluated in the selection process.

2. Evaluate the Instruments: Select the Step 2 tab for each identified instrument from the Step 1 Instrument Analyzing process to evaluate the instruments using the Instrument Evaluating Questions. The Step 2 tab for each instrument listed in alphabetical order, includes the Individual Instrument Reports, and the Instrument Evaluating Questions for the selection committee to review. You may also refer to the Descriptions and Terms tab as needed.

Descriptions and Terms

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."

Content Features: Depth of Coverage Scoring Rubric

Very Strong	Very strongly addresses key aspects of the skill
Strong	Strongly addresses key aspects of the skill
Moderate	Moderately addresses key aspects of the skill
Minimal	Minimally addresses key aspects of the skill
Not at all	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 Progress Monitoring Instruments are presented in bold. 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.

Psychometrics

Reliability is the consistency with which scores on an 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. Reliability is the consistency with which scores on a measurement instrument measure an underlying construct.

- Not all aspects of reliability will be applicable to all Instruments.
- 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 instrument 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 instrument

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 instrument

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 instrument

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 instrument

Validity refers to how well the components of an instrument measure what they are intended to measure. This review focuses on construct validity (i.e., concurrent, convergent, discriminative) and predictive 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 instrument

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 instrument

Generalizability is the degree to which the results can be applied to a broader context. Thus, for this to happen, the sample being tested should accurately represent the broader population. This is the degree to which the sample(s) of students used to develop the instrument and establish psychometric properties is sufficiently large and demographically similar to the Texas student population.

For each sample employed, score in the following manner:

Representativeness
3 = Representative 2 = Relatively representative 1 = Not very representative 0 = Not provided

Sample Size
3 = Large 2 = Moderate 1 = Limited 0 = Not provided

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

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

Growth/Improvement: Growth or Improvement (a.k.a. responsiveness or sensitivity to change) refers to the ability instruments have to measure meaningful differences in the constructs of interest. An instrument is said to be sensitive to growth when it measures increases and decreases in the construct measured such that scores increase with skill improvement and decrease when skills decline.

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

Administrative Feasibility

Feasibility was rated holistically in regard to teacher feasibility, student feasibility, and administrator feasibility.

Feasibility Feature	Components to Consider	Rating Description
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. • Instrument is visually appealing. • Instrument 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