Guidance for Evaluating Locally Selected Measures for Use in the Future Ready PA Index
Grade 3 Reading and Grade 7 Mathematics

Introduction

Two of the several new indicators identified for inclusion in the Future Ready PA Index (Index) are indicators of "on-track" performance in Grade 3 Reading and Grade 7 Mathematics. These indicators were selected for inclusion considering that research suggests a positive relationship between student attainment of core knowledge and skills in these grades and content areas and future academic success. In addition, feedback gathered from stakeholders across the state indicated that reporting student performance in these areas would be a useful and important contribution to the dashboard of school-specific information provided by the Index.

Purpose

The purpose of this document is to provide guidance for school entities to use when identifying and approving a locally selected assessment for the On-Track indicator included in the Index. It is the responsibility of LEA leadership to evaluate the quality and sufficiency of evidence provided by the vendor to support the use of the assessment and interpretation of results.
Ultimately, school entities will be responsible for determining whether the evidence provided meets the requirements for relevance and quality defined within this document.

For schools that elect not to use a locally selected assessment, the reporting for the On-Track indicator will default to the corresponding PSSA Grade 3 ELA or Grade 7 Mathematics score. Please note that regardless of whether a locally selected assessment is employed for the On-Track indicator, *ELA and Math PSSA scores will be required for the State Assessment Proficiency and Growth Measures* in the Future Ready PA Index. There is no local option to substitute testing for accountability purposes.

Implementation

Step 1: Initial Determination of Eligibility by the School Entity

The initial goal is to determine if the use of the assessment meets the minimum, non-technical requirements necessary to ensure fairness and consistency in administration, scoring, and reporting. Table 1 below addresses these requirements for the assessment.

Table 1 **Checklist of Minimum Requirements for a Locally Selected Assessment** **Grade 3 Reading or Grade 7 Mathematics Indicator** *To be completed and retained by the LEA*		
Requirements	**YES**	**NO**
The vendor provides clear test administration guidelines that serve to ensure consistency in the test administration process		
If teachers score items, there are guidelines, training materials and scoring rubrics that support fairness and consistency in the scoring process.		
Multiple forms and/or versions are provided. A single form or version of the test is not administered multiple times throughout the year (e.g., for practice, pre-test, benchmark, and summative purposes).		
The assessment has a cut score or performance level that represents expected or on-track performance within the grade-level[1]		
The assessment provides for student-level scores/results that can be aggregated and reported at the school level (i.e., not simply a narrative reflecting performance).		
The assessment provides for accessibility features or accommodations that allow it to be used by a minimum of 95% of the third (or seventh) grade students in a given school who are enrolled for a full academic year.		
The vendor provides support regarding the administration of the assessment for students who require accommodations.		
The school entity has procedures and resources in place to ensure data and results are collected and aggregated correctly and can be entered into PIMS to support Index reporting.		
Teachers/test administrators use and follow the vendor-provided test administration guidelines.		
The test must be administered at a similar point in time for all students within the school. (Either the test is given to all students within a specified testing window or the point in the instructional sequence within which a test should be administered is clearly specified.)		
If teachers are responsible for scoring all or certain components of the assessment, they receive appropriate training to do so fairly and consistently across all students. Materials and procedures are in place to ensure scoring procedures are conducted with fidelity.		
Students will receive appropriate accommodations based upon documented need (e.g., IEP, LIEP, 504 plan).		

If the school entity determines an assessment satisfactorily meets the Step 1 minimum requirements, the school entity proceeds to Step 2, which will be supplied by the manufacturer or vendor.

If the school entity determines an assessment does not meet the minimum requirements, there is no reason to proceed, the assessment is not appropriate for the intended use.

Step 2: Assessment Evaluation – Collect Required Evidence from Vendor to Support Locally Selected Assessment Eligibility

If an assessment meets all of the Step 1 requirements, the school entity should proceed to ask the vendor to establish a technical statement supporting the use of the selected test (or some component of it) as a Grade 3 Reading/Grade 7 Mathematics Success indicator. The test vendor should provide evidence to support each of the statements below:

1. The content and skills addressed by the test are relevant and sufficient for making decisions about whether students are on track relative to the specific academic content addressed.
2. Items, tasks, and stimuli are aligned to the skills identified as the focus of the assessment.
3. The assessment is fair for all students in the intended test taking population.
4. The performance standard (i.e., cut score/measurement) supports inferences about whether a student is on track for future success.
5. Assessment results are precise enough to support consistent decisions about whether a student is on track relative to academic content.
6. Test development, administration, and scoring procedures provide for comparable test results between students and over test adminstrations/years.
7. Assessment results are related to other measures considered indicative of future success.
8. Score reports are useful and easy to interpret.

For each statement, the vendor should briefly describe how procedures, results, and research associated with the assessment support the specified statement or claim. The argument should be coherent and easy to understand. For example, in support of Statement 4 related to performance standards, the vendor should provide an argument supporting the appropriateness of the established cut score for making decisions regarding whether a student is “on-track” to future success. The argument should briefly summarize the procedures, data, and materials used to support the claims.

A summary of the questions that should be addressed for each statement and the criteria by which each argument should be evaluated is provided in the Locally Selected Assessment Evaluation Matrix below.

If/when the vendor provides the requested information, school entities should proceed to and complete Step 3. The school entity must retain the technical statement provided by the vendor.

If the vendor does not agree to provide the requested information or suggests that the assessment is not appropriate for the intended use, school entities should not proceed to Step 3.

Table 2
Locally Selected Assessment Evaluation Matrix
Grade 3 Reading or Grade 7 Mathematics Indicator
To be used by a manufacturer or vendor to provide a technical statement

Statement to be Supported	Core Questions Driving the Collection of Evidence	The argument supporting this statement should draw upon sources of evidence such as the following:
1. The content and skills addressed by the test are relevant and sufficient for making decisions about whether students are on track for future success.	For what purpose was this assessment developed? How are results intended to be used? What is the content domain addressed by the test and by what process was it defined? What research supports the use of this assessment as a means of evaluating whether a student is on-track to future success? How and in what way does the test blueprint address those skills identified by research as necessary for future success?	• Documentation summarizing the primary purpose of the assessment and the process/participants used to define the content domain considering that purpose. • A test blueprint or test specifications document which clearly outlines the primary knowledge and skills targeted for assessment and how they will be measured. • Documentation showing the relationship between the type and range of evidence collected by the assessment and those skills/competencies deemed important for future success in the grade/content area (i.e., as documented in articles, research, frameworks that indicate those skills which evidence shows are related to future success).
3. Items, tasks, and stimuli are aligned to the skills identified as the focus of the assessment.	What procedures are in place to ensure that test items demonstrate the intended knowledge, skills, and competencies? To what extent does a given assessment	• Description of the item/task development and review process. • Tables/documents indicating the type of evidence expected to reflect student understanding of the skills identified for assessment. • Item and task development specifications that indicate the features necessary to elicit the type of evidence desired.

Table 2 Locally Selected Assessment Evaluation Matrix Grade 3 Reading or Grade 7 Mathematics Indicator *To be used by a manufacturer or vendor to provide a technical statement*		
Statement to be Supported	**Core Questions Driving the Collection of Evidence**	**The argument supporting this statement should draw upon sources of evidence such as the following:**
	address the full range of content and skills defined within the test blueprint? Are the text passages for reading assessments of appropriate length and complexity for this course?	• Passage specifications and procedures for determining and evaluating text complexity (when appropriate). • Descriptions of any studies conducted to ensure items/tasks measure the skills intended (e.g., usability analyses). • Tables reflecting the cognitive demand of the assessment items/tasks relative to the cognitive demand of the standards.
4. The assessment is fair for all students in the intended test taking population.	For what population of students was this assessment developed? What procedures are in place to ensure test items and tasks are appropriate for all students in the test taking population? Do provided accessibility features and accommodations support all students in the intended test taking population?	• Documentation of the intended test taking population, and those students for which the assessment may not be appropriate. • Descriptions of item, task, stimuli and test development and review procedures used to ensure the accessibility and fairness of items and tasks (e.g., including bias and sensitivity review procedures and analyses). • Item and task development specifications. • Results from item tryouts and/or usability analyses. Test Administration Guidelines • Lists of accommodations/access features provided by the test and to whom they are intended to serve. • White papers on defining accessibility for the program, where appropriate.
5. The performance standard (i.e., cut score/ measurement) supports inferences about whether a student is on track for future success.	What procedure was used to establish the cut-score used to represent "on-track" performance?	• Procedures used to develop any performance level descriptors (PLDs), or skill-based expectations for student achievement at the benchmark (if appropriate). • Detailed summary of the standard setting process, including who was involved, when it occurred and

Table 2 Locally Selected Assessment Evaluation Matrix Grade 3 Reading or Grade 7 Mathematics Indicator *To be used by a manufacturer or vendor to provide a technical statement*		
Statement to be Supported	**Core Questions Driving the Collection of Evidence**	**The argument supporting this statement should draw upon sources of evidence such as the following:**
	How does that procedure support the use of the cut score as an indicator of whether a student is on track to future success? What research supports the use of the performance standard/cut score for making accurate inference about future success?	any external data used to inform the specification of cuts. • Descriptions of any studies conducted to inform the standard setting process. • Results of studies evaluating the appropriateness/accuracy of the cut score for identifying those who are likely to succeed.
6. Assessment results are precise enough to support consistent decisions about whether a student is on-track.	How is the reliability of scores evaluated? What procedures are in place to ensure that test scores and performance classifications are not significantly influenced by error (i.e., factors that interfere with making accurate decisions about students)?	• Results of reliability analyses and interpretations of provided results (e.g., reliability coefficients, classification accuracy, decision consistency). • Procedures and guidelines are used to ensure consistency in human scoring (when appropriate). • Rater agreement analyses. • Quality control procedures related to scoring of selected response items. ○ Independent key verification ○ Statistical analysis of items
7. Test development, administration, and scoring procedures provide for comparable test results between students and over years.	What procedures are in place to ensure that the test produces comparable results across students and from one assessment	• Test development specifications and review procedures. • Test administration guidelines and procedures. • Scoring procedures. • Scaling and equating procedures and results.

Table 2 Locally Selected Assessment Evaluation Matrix Grade 3 Reading or Grade 7 Mathematics Indicator *To be used by a manufacturer or vendor to provide a technical statement*		
Statement to be Supported	**Core Questions Driving the Collection of Evidence**	**The argument supporting this statement should draw upon sources of evidence such as the following:**
	administration to the next? Does data suggest these procedures are working as intended?	• Rater drift analyses.
8. Assessment results are related to other measures considered indicative of future success.	Is there a relationship between the assessment results and other external variables indicative of success? To what extent do assessment results confirm or deny what other indicators measuring related skills are telling me?	• Research studies. • Correlations between performance on the assessment and other measures in the content domain. • Relationship between student performance classifications (e.g., on-track/not on-track) and measures indicative of success.
9. Scoring reports are useful and easy to interpret.	Can stakeholders easily locate information that helps them understand what the test measures and how Index results are to be interpreted?	• Score Interpretation Guides. • Test Blueprints or a summary of the assessed content domain. • Sample Test Items. • Scoring Rubrics (if appropriate).

Step 3: Evaluation of the Quality and Sufficiency of the Provided Evidence by the School Entity

Once the vendor has supplied a response to each of the statements outlined in Table 2, it is the responsibility of the school entity to review and evaluate the sufficiency of those responses. The overarching goal of the evaluation process is to determine whether the vendor has provided sufficient evidence supporting the use of the test for the Index indicator. Guidance supporting this process, including features characterizing a high-quality response and expected evidence, are provided in Table 3: Evaluation Summary Document. In Table 3 there are ten categories of evidence. Ultimately, it will be school entity's decision as to whether the assessment provides for high-quality, useful results.

When evaluating evidence provided by vendor, LEA leadership should be able to conclude with confidence that the evidence for each of the ten categories is acceptable and is marked *Yes*. For each category, school entities should evaluate the evidence and draw a conclusion as to whether the intent of the specific category was met. If the school entity finds an area of weakness, additional information can be requested from the vendor to support that area of concern and/or it should be documented in the summary report.

Evaluation expertise may come from multiple sources, e.g., an assessment director, local educators, a local university, or private contractor. When possible, it is also strongly encouraged to include evaluators with experience teaching/evaluating students with disabilities and English learners (to evaluate evidence regarding fairness and inclusion) and someone with a background in assessment development or measurement theory, who can review technical evidence, as needed, to inform the overall evaluation.

Upon completion of the evaluation process through Table 3, the school entity will generate a summary document to include the names/qualifications of the participants, a brief summary of the review process, and the overall decision made by the evaluator(s). The summary of the review process should include statements that document the quality of the evidence reviewed for each category and how that evidence supports the decision to use the assessment.

While the summary document is not intended for public distribution, it could ultimately be provided in response to external requests regarding the school entity's rationale for recommending the assessment and therefore should be written accordingly.

Once the review process is complete, the school entity should make a holistic decision as to whether the information provided supports the use of the assessment as an on-track indicator. The superintendent/chief school administrator acknowledges that the assessment meets the evaluation criteria by signing the assurances included with the Accuracy Certification Statement provided during PIMS reporting.

Table 3 Evaluation Summary Document Grade 3 Reading or Grade 7 Mathematics Indicator *To be completed and retained by the LEA*			
LEA:	**Check one:** ☐ **Grade 3 Reading** ☐ **Grade 7 Mathematics**		
Names/Qualifications of LEA Evaluators:			
Summary of Review Process:			
Overall Decision:			
The Evidence Provided in Step 2 Demonstrates that…			
Category 1: **The content and skills addressed by the test are relevant and sufficient for making decisions about whether students are on track for future success.**		**YES**	**NO**
• An explanation is provided describing how the purpose for/uses of the test were designed to support goals of the Grade 3 Reading/Grade 7 Math on-track indicator. • The researched relationship between the skills measured on the test and those necessary for/related to future success is documented. • Evidence is provided supporting the breadth and depth of the assessed content domain in making useful inferences about whether students are on track.			
Statements Documenting Quality of Evidence:			
Category 2: **Items, tasks and stimuli are aligned to the skills identified as the focus of the assessment.**		**YES**	**NO**
• Test development procedures include mechanisms for evaluating and trying out items/tasks prior to use. • Detailed item/task development process/descriptions are included. • If passages are required to support assessment, procedures and specifications are in place to support the evaluation of text complexity to ensure it is appropriate for the grade level/content area.			
Statements Documenting Quality of Evidence:			
Category 3: **The assessment is fair for all students in the intended test taking population and e*ffort was made to ensure an equitable opportunity to all students in the intended test taking population.***		**YES**	**NO**

- Item, task, and/or test development specifications are written to ensure maximum readability, comprehensibility, and legibility.
- All items are reviewed for bias/sensitivity fairness prior to use.
- Analyses are conducted so that suspect items, tasks, or stimuli can be flagged for review prior to use.
- Lists of accommodations/access features are provided by the test and to whom they are intended to serve.
- The intended test taking population is clearly documented in conjunction with a list of those accommodations/accessibility features available to support participation.
- Educators are provided with guidance to help determine which types of features/accommodations should be used for different students.
- Test administration guidelines ensure all students are provided with a fair, appropriate testing environment.

Statements Documenting Quality of Evidence:

Category 4: **The performance standard (i.e., cut score/measurement) supports inferences about whether a student is on track for future success.**	**YES**	**NO**

- A detailed description of the process/technique used to establish the performance standard is provided.
- If performance level descriptors were used to support standard setting, content experts were involved in the recommendation and/or review of the cut scores and technical experts supported the development and facilitation of the standard setting process.
- If standards are based solely on the relationship between the test and some other external measure (e.g., ACT, state test), the rationale for using that measure as the basis for the cut is described.
- Data and materials used during the standard setting that suggest the cut-score reflects end of grade or on-track expectations, such as:
  - Performance level descriptors that reflect grade-level expectations.
  - Use of performance/results from other assessments measuring similar skills, constructs, and/or content domains.
  - The following year's grades in school or graduation from high school.

Statements Documenting Quality of Evidence:

Category 5: **Assessment results are precise enough to support consistent decisions about whether a student is on track.**	**YES**	**NO**

- Results of reliability analyses were provided (e.g., reliability coefficients, classification accuracy, decision consistency).
- Procedures and guidelines are in place to ensure consistency in human scoring (when appropriate).
- Inter-rater reliability data were provided.
- Quality control procedures are documented re: scoring of selected response items.
  - Independent key verification
  - Statistical analysis of items

Statements Documenting Quality of Evidence:

Category 6: **Test development, administration, and scoring procedures provide for comparable test results between students and over years.**	**YES**	**NO**

- Item design/development materials are written at a level of detail that supports consistency in the development of items over time.
- Test specifications are written to support equivalence in the development of test forms, in terms of content representation and overall text complexity.
- Review procedures are in place to evaluate the equivalence of test forms prior to operational use.
- Test administration procedures are detailed enough to support standardization across forms, sites, and administrations.
- Scoring rubrics are reviewed and piloted for clarity and utility prior to operational use.
- Scoring rubrics provide exemplar responses to ensure consistency in the scoring process.

Statements Documenting Quality of Evidence:

Category 7: Assessment results are related to other measures considered indicative of future success.	**YES**	**NO**

- Results of research studies reflect a clear relationship between attainment of the benchmark and other measures indicative of being on-track for success in the content domain, such as:
  - Performance on other assessments measuring similar skills/constructs.
  - The following year's grades in school
  - Graduation from high school
- Procedures are in place to evaluate the appropriateness of the cut score over time and make modifications, if deemed necessary.

Statements Documenting Quality of Evidence:

Category 8: Score reports are useful and easy to interpret.	**YES**	**NO**

- User-friendly documents and resources (e.g., score interpretation guides) are available to help parents, students, and other stakeholders understand what the test measures and what it means to be on-track.
- Sample test items (and student responses) are provided to illustrate expectations for students.

Statements Documenting Quality of Evidence:

Data Reporting and Monitoring

School entities using a local assessment are responsible for reporting/uploading individual student data into PIMS to report on-track performance of each individual student in Grade 3 Reading and/or Grade 7 Mathematics.

- In Row 19 of the **Table: STUDENT_LOCAL_ASSMNT_SUBTEST,** an indication of whether the student met the standard identified, On-track Proficiency, is required. Valid value: Y or N only.
    - Enter “Y” in Field 19 of the template if the student meets the identified standard.
    - Enter “N” in Field 19 of the template if the student does not meet the identified standard

By signing the assurances included with the Accuracy Certification Statement provided during PIMS reporting, the superintendent/chief school administrator verifies the quality of the locally selected assessment and the accuracy of the data reported.

1. If an alternate assessment is chosen, all students reported by the school entity must use the same assessment, or all students will default to PSSA.
2. All students in the same grade level throughout all buildings in the LEA must use the alternate assessment, or all students will default to the PSSA.

School entities are also reminded that the assessment must provide for accessibility features or accommodations that allow it to be used by a minimum of 95% of the students in each school who are enrolled for a full academic year. Schools who do not attain a 95% participation rate on the locally selected assessment will have their data reverted to PSSA assessment data. Overall, it is critical to verify that every effort was made to ensure an equitable opportunity to all students in the intended test taking population.

During PSSA/Keystone monitoring, monitors may request to see all documentation used to support the decision by the school entity that the locally selected assessment is appropriate as outlined in Steps 1-3 of this document. Monitors *may* also request to see the assessment or sample assessment, samples of student work, records of student data, and/or evidence of the methodology used to determine cut scores.