



Universal Screener Panel Report and Approved List

Division of Instructional Programs, Literacy Programs and Initiatives

February 2026

MARYLAND STATE DEPARTMENT OF EDUCATION

Carey M. Wright, Ed.D.

State Superintendent of Schools

Tenette Y. Smith, Ed.D

Chief Academic officer

Office of Teaching and Learning

Wes Moore

Governor

MARYLAND STATE BOARD OF EDUCATION

Joshua L. Michael, Ph.D., President

Monica Goldson, Ed.D., Vice President

O'Marie Barnes, Student Member

James C. Bell, Jr., Ed.D.

Chuen-Chin Bianca Chang, MSN, PNP, RN-BC

Alverne "Chet" Chesterfield

Kenny Clash

Nick Greer

Dr. Irma E. Johnson

Marcy Leonard

Dr. Kim Lewis

Xiomara V. Medina, M.Ed.

Samir Paul, Esq.

Contents

Purpose	3
Universal Screener Review Panel Overview	4
Universal Screener Review Panel.....	5
Collective Role	6
Universal Screener Review Outcomes	7
Recommended Screeners	8
Not Recommended Screeners.....	9
Approved Universal Screeners.....	12
Required Measures Per COMAR 13A.03.08.....	13
Appendix A: Universal Screener Review Panel Rubric	14
Introduction.....	15
Description of Assessment Battery	15
Appendix B: Recruitment and Review Timeline	26
Appendix C: Definitions.....	27
Definitions.....	28
Appendix D: Renaissance Scoring	29

Purpose

The purpose of this report is to share the Maryland State Department of Education's (MSDE), Universal Screener Panel's vetted and approved universal screeners for use by local education agencies (LEAs) in Pre-Kindergarten through grade 3. These approved screeners support the implementation of Maryland's [Comprehensive PreK-3 Literacy Policy](#) and fulfill requirements outlined in the [Ready to Read Act](#), and [COMAR 13A.03.08, Students at Risk for Reading Difficulties](#). Approved universal reading screeners are administered three times each school year, at the beginning, middle, and end of year, to provide consistent and timely information about students' early reading development.

The purpose of universal reading screeners is to identify students who may be at-risk for reading difficulties as early as possible, including students who may demonstrate risk of dyslexia. Approved screeners measure key foundational reading skills aligned to the science of reading, including phonological and phonemic awareness, sound-symbol recognition, decoding, fluency, and rapid automatic naming (RAN). Some screeners may also include additional measures, such as encoding or reading comprehension.

Early childhood is a critical period for literacy development, and research consistently shows that early identification followed by targeted instruction and intervention leads to stronger reading outcomes. Universal screening ensures that all students are monitored regularly so that instructional supports can be provided early, before reading challenges become more significant.

Screening results are used to inform instructional planning and decision-making. When screening data indicates a potential reading difficulty, educators use additional diagnostic tools and ongoing progress monitoring to better understand student needs and to adjust instruction accordingly. Approved screeners are selected to minimize instructional disruption, support effective use of data by educators, and provide timely results to teachers, school leaders, and families.

Through the use of the Universal Screener Panel's approved universal reading screeners, Maryland schools strengthen early literacy instruction, promote excellent and equitable access to supports, and help ensure that all students develop the foundational reading skills necessary for long-term academic success.

Universal Screener Review Panel Overview

The Universal Screener Review Panel evaluated the submitted literacy screening instruments to determine their appropriateness for identifying students at-risk for reading difficulties, in accordance with COMAR 13A.03.08, MSDE's Comprehensive PreK-3 Literacy Policy, and the Ready to Read Act. The panel's work ensured that approved screeners were developmentally appropriate, evidence-based, and suitable for Maryland's diverse student population.

The panel reviewed the literacy constructs measured at each grade level, including the required foundational reading skills named in the Comprehensive PreK-3 Literacy Policy: phonological and phonemic awareness, sound-symbol recognition, decoding, fluency, and rapid automatic naming. The panel devoted attention to theoretical grounding and developmental appropriateness of each universal screening instrument.

The review also examined administration and scoring procedures, including intended use, assessor qualifications, accessibility for Multilingual Learners and students with disabilities, and clarity of score interpretation. Instruments were evaluated for guidance on English language proficiency thresholds and appropriate use of alternate language assessments.

A core focus of the panel's work was the evaluation of psychometric quality, including reliability, validity, classification accuracy, and decision rules, with evidence reviewed overall and across student subgroups representative of Maryland.

Finally, the panel assessed the extent to which screeners provided actionable guidance and resources for educators and families, including professional learning supports, instructional guidance, clear reporting, multilingual resources, and privacy protections. Using the Universal Screener Review Panel Rubric (Appendix A), the panel rated evidence across criteria and produced descriptions of approved screeners and their valid uses to inform local education agency decision-making.

Universal Screener Review Panel

The Universal Screener Review Panel was composed of state and district leaders with deep expertise in literacy instruction, assessment, accountability, and psychometrics, ensuring that decisions reflected statewide priorities as well as district-level implementation realities.

Panelist Name	Role	LEA/Agency
Chelsea Brewer (Panel Chairperson)	Executive Director, Literacy Programs and Initiatives	Maryland State Department of Education
Jaime Bowers	Director of Testing, Research and Evaluation	Prince George's County Public Schools
Susan Copen	Coordinator of Elementary Integrated Literacy	Frederick County Public Schools
Paul Katula	Scoring Specialist	Maryland State Department of Education
Meg Kinlaw	Coordinator, Literacy Programs and Initiatives	Maryland State Department of Education
Erika McCrea	Executive Director, State Assessments	Maryland State Department of Education
Phillip Snyder	Supervisor of Accountability	Harford County Public Schools
Suzanne Tirado	Coordinator, Literacy Programs and Initiatives	Maryland State Department of Education
Mei-Hui Wang	Psychometrician, Office of Research	Anne Arundel County Public Schools
Stephanie Weager	English Language Arts/Literacy Assessment Project Manager	Maryland State Department of Education

COLLECTIVE ROLE

Together, the panel reviewed universal screener tools, administration and analysis guidance, and data to:

- Evaluate the effectiveness and appropriateness of screening tools
- Ensure fair, valid, and reliable interpretation of results
- Support data-informed decisions related to literacy instruction and intervention
- Align screening practices with statewide literacy goals and accountability expectations

Universal Screener Review Outcomes

RECOMMENDED SCREENERS

1. **Amira ISIP Assess**

The Amira Learning Universal Literacy Screener (ISIP Assess) is an AI-powered, Science of Reading aligned assessment system designed for K–3 universal screening, progress monitoring, and dyslexia risk identification. Administered primarily in a fully virtual format, the assessment requires students to use headphones in a quiet setting, with adult monitoring recommended to ensure completion. Small group or one-to-one administration is most appropriate. Each administration lasts approximately 12–20 minutes. The assessment is adaptive, using Item Response Theory (IRT), and evaluates key literacy domains including phonological and phonemic awareness, sound–symbol recognition, decoding, fluency, and Rapid Automatized Naming (RAN). A paper-based option is available for students who require accommodations.

Amira is available in English and Spanish and was developed with bilingual literacy experts. Reliability evidence is strong, with coefficients exceeding 0.80 across internal consistency, test–retest, and inter-rater measures, and validity evidence, including sensitivity and specificity for early reading risk, is well documented. AI-supported administration and scoring eliminate the need for teacher scoring calibration and generate composite scores that color code students by dyslexia risk level. The platform includes educator dashboards, family reports, accessibility supports, and data security protections, and has received strong external validation, including the high ratings from the National Center on Intensive Intervention (NCII). Overall, the panel recommends Amira as a universal literacy screener.

2. **Amplify mCLASS with DIBELS 8th Edition**

Amplify mCLASS with DIBELS 8th Edition is a universal literacy screening system with a strong research base, designed to identify students at-risk for reading difficulties, including dyslexia, and to support progress monitoring and instructional decision-making. The assessment is teacher administered one-to-one, with the teacher recording student responses using a tablet or computer while the student utilizes a paper booklet or digital stimuli. Administration typically takes 12–20 minutes per student. The system is consistent with International Dyslexia Association (IDA) recommendations, incorporating measures such as Letter Naming Fluency (LNF), Phonemic Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Word Reading Fluency (WRF), Oral Reading Fluency (ORF), and Rapid Automatized Naming (RAN) using letter names (RAN).

DIBELS 8 demonstrates strong reliability and validity, with multiple forms of reliability reported transparently and coefficients generally meeting or exceeding .80. Concurrent and predictive validity are supported through moderate-to-strong correlations with established measures, including predictive validity coefficients ranging from .68 to .85. mClass supports Spanish administration through mCLASS Lectura and provides standardized accommodations to ensure valid scoring. Family communication is supported through mCLASS Home Connect, which offers family friendly reports and activities in English and Spanish. Amplify provides extensive professional learning, coaching, and implementation support, and maintains strong data privacy and security practices aligned with recognized frameworks. The panel reached consensus to recommend mCLASS with DIBELS 8th Edition as a universal literacy screener.

3. **Renaissance STAR Reading**

Renaissance STAR Reading (including STAR Early Literacy) is a comprehensive universal screening system for grades PreK-3, with STAR Early Literacy serving as a component of MSDE's Kindergarten Readiness Assessment (KRA). The assessment uses a computer-adaptive design to efficiently measure phonological and phonemic awareness, sound-symbol recognition, and decoding, with an average administration time of approximately nine minutes per student in grades K-3. To meet the MSDE Universal Screening criteria, the Star Curriculum Based Measures (CBM) in fluency and RAN must also be administered, adding an additional five minutes. Both computer-based and paper-based options are available. An authentic Spanish-language version is offered.

STAR Literacy demonstrates sufficient evidence of reliability, validity, and usability to meet state requirements for universal screening. Evidence indicates moderate to strong technical adequacy, with reported reliability and validity estimates typically ranging from the mid-0.70s to low-0.80s, and moderate sensitivity and specificity for identifying students at-risk for reading difficulties. Validation studies include diverse student samples across grade levels and language backgrounds, including English and Spanish speakers. Overall construct alignment reflects developmental appropriateness, and administration and scoring procedures are clearly articulated and research informed. The system provides actionable, automatic reporting at the student and classroom levels, supported by professional learning resources and educator tools designed to guide instructional decision-making. The panel recommends Renaissance STAR Reading as a Universal Screener.

NOT RECOMMENDED SCREENERS

1. **Curriculum Associates: i-Ready**

The i-Ready assessment system from Curriculum Associates includes a computer-based adaptive diagnostic and teacher-led literacy tasks, both of which would need to be administered for the tool to function as a universal screener under MSDE requirements. Strengths of the system include strong reliability and validity evidence, predictive relationships with MCAP performance, and the inclusion of intervention and progress-monitoring tools housed within a single administrative platform. An authentically written Spanish version is available, along with additional one-to-one literacy tasks for deeper diagnostic follow-up.

However, the panel identified several significant concerns that ultimately led to a decision not to recommend the tool. The Comprehensive PreK-3 Literacy Policy states that MSDE shall consider “the time required to conduct the assessment, with the intention of minimizing the impact on instructional time.” Substantial time is required for i-Ready administration: approximately 45 minutes for kindergarten and grade 1 and up to 80 minutes for upper grades for the diagnostic alone. This does not include the required teacher-led literacy tasks that qualify i-Ready as a universal screener. The assessment also requires computer-based testing conditions (headphones, small-group or 1:1 administration, and adult monitoring), and some students may need multiple sittings to complete it, further reducing instructional time. Additionally, there are no low-tech or paper-based options for the diagnostic component, including the Spanish version, which would require districts to identify alternative screening tools for students with paper-pencil accommodations on an Individualized Education Plan (IEP) or 504 plan. Taken together, the breadth, administration demands, and time requirements for i-Ready, does not meet the requirements of the Comprehensive PreK-3 Policy, leading the committee to conclude that it does not meet the criteria for recommendation.

4. **DFE Professional**

The DFE Professional Universal Literacy Screener is an early literacy assessment tool focused on foundational reading skills and educator-friendly administration and scoring procedures. However, the review panel identified significant gaps in the evidence required to support full approval. Although the request for information references peer-reviewed research and general psychometric thresholds, the submission lacks sufficient technical documentation to verify claims related to reliability, validity, and classification accuracy. Detailed data tables, subgroup analyses, and clear evidence of performance across diverse student populations were not provided, limiting confidence in the tool’s ability to accurately identify students at-risk.

Additional concerns were raised regarding instructional follow-through and communication supports. While procedural guidance for educators is clear, there is minimal direction on how assessment results should translate into instructional decisions or interventions. Family communication and multilingual accessibility are acknowledged conceptually, but evidence of family-facing materials, culturally responsive resources, or structured communication frameworks is limited.

Overall, the DFE Professional screener appears thoughtfully designed and usable, but insufficient evidence was provided to substantiate its technical quality and equitable implementation. As a result, the committee did not approve this vendor, citing the need for

stronger psychometric documentation, clearer language proficiency guidance, and more robust family communication supports before reconsideration.

5. **EPS Learning Amira**

The Amira assessment presented by EPS Learning is a PK-3, automated, AI-driven tool designed to function as a three-in-one system for benchmarking, progress monitoring, and dyslexia screening. The proposal highlighted several strengths, including the efficiency of automated administration, the integration of multiple screening purposes within a single platform, and the overall quality of the underlying assessment, which the committee recognizes as strong. However, the assessment was not recommended because it represents a repackaging or resale of Amira, a product that has already been reviewed and approved, and the proposal included incomplete information. As a result, the committee determined there was no need to approve the same tool through another provider.

6. **Houghton Mifflin Harcourt: MAP Reading Fluency**

MAP Reading Fluency by Houghton Mifflin Harcourt is a digital, computer-adaptive assessment for students in grades PreK-5. It includes a universal screener, benchmark assessments, progress monitoring, an optional dyslexia screener, and an optional virtual tutoring add-on. MAP features an engaging design and speech-scoring technology for oral reading. The assessment also the inclusion of linguistically authentic Spanish language assessment.

While some evidence of validity is reported, there were uneven psychometric results across domains, raising concerns about consistency across the assessment. Although reliability and validity for Oral Reading Fluency are strong and show high correlations with established measures, evidence for other foundational skills is weaker, and language comprehension demonstrates limited validity. In addition, the assessment does not define minimum language proficiency thresholds for valid score interpretation, and documentation related to classification accuracy, subgroup-level decision rules, and the representativeness of validation samples relative to Maryland populations is insufficient for use as a universal screener.

7. **MindPlay**

MindPlay is a computer-based, adaptive universal screener designed for students in grades K–12. The assessment takes approximately 30 minutes to administer, with an additional optional 30-minute dyslexia screener. Strengths of the tool include compatibility with multiple data platforms, adaptive technology features, and limited language supports.

However, the assessment was not recommended due to significant concerns about its alignment with the Comprehensive PreK-3 Literacy Policy's universal screener guidelines. The amount of time required, and the number of skills assessed would negatively impact on instructional time, particularly given the purpose of a brief, high level screener. Theoretical foundations and the rationale for assessment design are not clearly explained, and claims related to norming for Multilingual Learners and students with disabilities are not supported with sufficient evidence. In addition, key psychometric information is missing, including defined cut scores, sensitivity, and specificity values, and details about the sample used for norming are limited. Other concerns include vague information about professional development, generic and insufficient support for Multilingual Learners (with the assessment available only in English), high-level descriptions of privacy and security practices without operational detail, and

the inclusion of subtests, such as grammar and visual scanning, that do not clearly align with the purpose of a universal screener. Taken together, these issues led the panel to decide not to recommend the MindPlay assessment.

8. NCS Pearson: aimswebPlus

NCS Pearson's aimswebPlus demonstrates a strong psychometric and operational foundation as an early literacy screening tool. The assessment is supported by recently updated norms based on large, representative samples, and reliability and validity evidence is well documented through detailed technical resources. The battery targets some key early literacy constructs across grades Pre-K through 3, with grade specific composites built from measures such as letter naming fluency, letter word sound fluency, oral reading fluency, vocabulary, and comprehension. Administration procedures and educator supports are robust, including extensive manuals, training resources, scoring guidance, and strong data privacy and confidentiality safeguards. The platform also offers Spanish language assessments and reports.

However, the assessment does not meet state screening requirements, which mandate the inclusion of phonological and phonemic awareness, sound-symbol recognition, decoding, fluency, and rapid automatic naming measures. Chief among these is the lack of a rapid automatic naming subtest, or an explicit phonics measure within the Grade 2–3 composite. The system restricts access to foundational skills assessment in grades 2–3 to those students who perform below expectations in the provided grade level measures. The kindergarten composite score is made up of only the letter naming fluency and letter word sounds fluency subtests, creating the potential for under identification of students needing supplemental instruction. Due to the missing subtests required, the panel did not recommend this assessment as a Universal Screener.

9. Stanford: Rapid Online Assessment of Reading (ROAR)

The ROAR assessment developed by Stanford presents an innovative approach to measuring literacy-related skills and demonstrates careful consideration in comparing its methods to other screening tools. The framework is organized into foundational skills, dyslexia screening, and language comprehension, and the proposal reflects thoughtful work in explaining how these components function together. The committee appreciated the exploration of alternative ways to assess language skills and noted potential promise for use with older students and in dual-language contexts.

However, the assessment was not recommended because it does not align with key requirements for a universal screener under the Universal Screening Panel Review Rubric. Most notably, ROAR does not meet COMAR 13A.03.08's definition of oral reading fluency, as students do not read aloud and fluency is assessed through silent reading. In addition, aspects of phonological awareness are limited to first-sound matching, last-sound matching, and rhyming, with no phoneme segmenting task, raising concerns about completeness and comparability with other screeners. The reporting structure relies on only three performance levels based on percentile ranks, which makes it difficult for districts to determine levels of risk, align results across assessments, and make consistent instructional decisions. Questions about how certain skills are assessed, what contributes to the composite score, and how "at-risk" would be defined equitably across districts further contributed to the decision not to recommend the ROAR assessment as Universal Screener.

Approved Universal Screeners

Vendor	Administration Time	Format	At – Risk identifier	Additional Language Options
Amira ISIP Assess	12-20 minutes	Computer Adaptive Test (CAT)	Amira Reading Master (ARM) Score: At-risk (Stronger signals/Red)	Available in Spanish: Amira Lectura Assess
Amplify mCLASS with DIBELS 8th edition	12-20 minutes	Curriculum Based Measure (CBM) 1:1 with student and teacher.	Composite Score: Well Below (Red)	Spanish version: mCLASS mLectura available
Renaissance STAR Reading	9-18 minutes	CAT with CBM 1:1 with student and teacher.	CAT: Below the 25 th percentile CBM: Below the 20 th percentile (see appendix D for details)	Spanish version: Star Lectura

*Screeners in other languages were not reviewed by the panel.

Required Measures Per COMAR 13A.03.08

Grade	Amira	Amplify DIBELS 8 th Edition	Renaissance Star Reading /Star Early Literacy
Kindergarten	Amira ISIP Assess	Letter Naming Fluency (LNF) Phonemic Segmentation Fluency (PSF) Nonsense Word Fluency (NWF) Word Reading Fluency (WRF) Rapid Automatized Naming (RAN)	Star Early Literacy Star Reading CBM: Rapid Automatic Naming
First Grade	Amira ISIP Assess	Letter Naming Fluency (LNF) Phonemic Segmentation Fluency (PSF) Nonsense Word Fluency (NWF) Word Reading Fluency (WRF) Oral Reading Fluency (ORF) Rapid Automatized Naming (RAN)	Star Early Literacy or Star Reading Star CBM Reading: Fluency, Rapid Automatic Naming
Second Grade	Amira ISIP Assess	Nonsense Word Fluency (NWF) Word Reading Fluency (WRF) Oral Reading Fluency (ORF) Maze Rapid Automatized Naming (RAN)	Star Early Literacy or Star Reading Star CBM Reading: Fluency, Rapid Automatic Naming
Third Grade	Amira ISIP Assess	Nonsense Word Fluency (NWF) Word Reading Fluency (WRF) Oral Reading Fluency (ORF) Maze Rapid Automatized Naming (RAN)	Star Early Literacy or Star Reading Star CBM Reading: Fluency, Rapid Automatic Naming

Appendix A: Universal Screener Review Panel Rubric

INTRODUCTION

The purpose of this rubric is to assist the Maryland State Department of Education (MSDE) and the PreK–3 Literacy Screener Review Panel in selecting high-quality universal screening instruments that accurately identify students at-risk of reading difficulties. This work is directed by Maryland's PreK–3 Comprehensive Literacy Policy, adopted by the State Board of Education in 2024, which establishes expectations for literacy instruction, universal screening, and early intervention. The rubric is further grounded in COMAR 13A.03.08, Students at Risk for Reading Difficulties, which defines regulatory requirements for screening, supplemental instruction, and reporting, and the Ready to Read Act of 2019, which requires early identification of reading difficulties and targeted supports for students.

The assessment battery under review should include developmentally appropriate constructs grounded in research and empirical evidence. Each instrument must consider a child's age, developmental level, background, and language acquisition, and must demonstrate strong psychometric properties. Evidence should reflect and respond to the needs of Maryland's diverse student population, including multilingual learners and students with disabilities.

This rubric is designed for use by the Maryland Screener Review Panel during the review and evaluation of submitted screening instruments. It aligns directly with the Review Elements and Evaluation Criteria and is organized into three major sections: Description of the Assessment Battery; Psychometrics; and Communication and Resources. Each section contains specific evaluation criteria, and each criterion is rated using three evidence levels: strong, moderate, and minimal. Collectively, the evaluation criteria provide a comprehensive framework for assessing the appropriateness of each instrument, rather than relying on any single criterion in isolation. Developmental appropriateness of the instruments is emphasized throughout the Description of Assessment Battery (Evaluation Criteria a, b, and c) and Psychometrics (Evaluation Criterion d) sections. Distinguishing features of each evidence level are determined based on both the quality of the evidence presented and the completeness of the supporting documentation.

To ensure consistency and objectivity, the Panel will present, discuss, and deliberate on the strengths and weaknesses of each screening instrument, including its overall appropriateness for accurately identifying students at-risk of reading difficulties. Following its review, the Panel will develop and publish a brief summary of information about each recommended and approved screening instrument. These summaries will be made available by MSDE to inform local education agencies and decision makers about the properties, affordances, and limitations of each instrument approved for use in Maryland.

DESCRIPTION OF ASSESSMENT BATTERY

1. Constructs and Rationale

The extent to which the screening instruments measure key constructs in a manner that is theoretically and empirically well-grounded (evaluated for each grade level and language represented)

- Theoretical frameworks and evidence, including their developmental appropriateness at each grade level, should be provided for the constructs/content in the screening instrument.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is a clear and compelling theoretical framework and evidence demonstrating the appropriateness of the constructs/content measured by the instrument for determining risk of reading difficulties at each grade level.	There is a reasonable theoretical framework and some evidence demonstrating the appropriateness of the constructs/content measured by the instrument for determining risk of reading difficulties at each grade level.	There is an unconvincing theoretical framework or limited evidence demonstrating the appropriateness of the constructs/content measured by the instrument are appropriate for determining risk for reading difficulties at each grade level.

- Constructs directly measured at each grade level should be listed and described. These may include, but are not limited to: oral language, phonological and phonemic awareness, decoding skills, letter-sound knowledge, knowledge of letter names, rapid automatized naming, visual attention, reading fluency, vocabulary, and language comprehension.

Strong Evidence	Moderate Evidence	Minimal Evidence
The content specifications - including the tasks and response modes for each task - are well supported by theoretical rationale and empirical evidence demonstrating their appropriateness for identifying risk of reading difficulties.	The content specifications - including the tasks and response modes for each task - are reasonably supported by theoretical rationale and empirical evidence demonstrating their appropriateness for identifying risk of reading difficulties.	The content specifications - including the tasks and response - modes for each task are poorly supported by theoretical rationale or empirical evidence demonstrating their appropriateness for identifying risk of reading difficulties.

- For instruments in languages other than English, the rationale and evidence for how constructs have been developed as appropriate to the language should also be provided,

Strong Evidence	Moderate Evidence	Minimal Evidence
There is a clear rationale and compelling evidence to support how the constructs/content have been developed for languages other than English.	There is a partial rationale and/or reasonable evidence to support how the constructs/content have been developed for languages other than English.	There is limited or no rationale or limited or unconvincing evidence to support how the constructs or content have been developed for languages other than English.

2. Administration and Scoring

The extent to which the mode of administration for the screening instruments is appropriate for the students being evaluated (by grade level and student need)

- Descriptions of tasks should include their intended and appropriate use.

Strong Evidence	Moderate Evidence	Minimal Evidence
Decisions about the appropriate use of tasks for different populations of students (e.g. by age, prior education, English learner status, disability status) are based on convincing evidence and are clearly described.	Decisions about the appropriate use of tasks for different populations of students (e.g. by age, prior education, English learner status, disability status) are based on at least partial evidence and are reasonably well described.	Decisions about the appropriate use of tasks for different populations of students (e.g. by age, prior education, English learner status, disability status) are based on weak or unclear evidence or are not described.

- For each task at each grade level, a description of the administration (individual or group administration; qualification of the assessor) and scoring format and platform, the number of items, assessment time, administration procedures and scoring procedures, and types of scores and their interpretation.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear guidance for appropriate administration that is well-grounded in evidence about how decisions are appropriate to the students being evaluated (by age and developmental level, language background, and student need), including consideration of all of the following: the mode of administration, assessor qualifications, guidelines for administration and the appropriate testing environment, and guidance for handling testing irregularities.	There is some guidance for administration that is grounded in partial evidence about how decisions are appropriate to the students being evaluated (by age and developmental level, language background, and student need), and including consideration of most of the following: the mode of administration, assessor qualifications, guidelines for administration and the appropriate testing environment, and guidance for handling testing irregularities.	There is unclear and/or limited guidance for appropriate administration, based on limited evidence about how decisions are appropriate to the students being evaluated (by age and developmental level, language background, and student need), and including consideration of some of the following: the mode of administration, assessor qualifications, guidelines for administration and the appropriate testing environment, and guidance for handling testing irregularities.
There is clear and detailed guidance for scoring the items/tasks and interpreting and reporting scores, considering students' age and developmental level, language backgrounds (e.g., those who speak English dialects and varieties), and any known disabilities (e.g., sight, speech or hearing).	There is some or reasonable guidance for scoring the items/tasks and interpreting and reporting scores, considering students' age and developmental level, language backgrounds (e.g., those who speak English dialects and varieties), and any known disabilities (e.g., speech or hearing).	There is unclear and/or limited guidance for scoring the items/tasks and interpreting and reporting scores, considering students' age and developmental level, language backgrounds (e.g., those who speak English dialects and varieties), and any known disabilities (e.g., speech or hearing).

- For instruments in languages other than English, how tasks have been constructed to appropriately reflect relevant language features.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is compelling evidence that supports how the tasks are constructed to appropriately reflect key relevant language features.	There is reasonable evidence that supports how the tasks are constructed to appropriately reflect relevant language features but does not adequately address all key relevant language features.	There is limited or unclear evidence that supports how the tasks are constructed to appropriately reflect key relevant language features.

3. Language Proficiency Thresholds

The extent to which the screening instruments offer well-grounded guidance for determining when a student has sufficient language proficiency for them to be appropriately used for the instrument to yield valid information in each assessed language should be provided.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear guidance, based on convincing evidence, about the minimum language proficiency level necessary for the instrument to yield valid information.	There is reasonable guidance about the minimum language proficiency level necessary for the instrument to yield valid information, but the evidence base is partial.	There is limited or little guidance about the minimum language proficiency necessary for the instrument to yield valid information or the evidence base is unclear or weak.

4. Psychometrics

Reliability and Validity

The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the Maryland student population

- Appropriate reliability estimates for different types of tasks (e.g., internal consistency, test–retest, alternate form, interrater agreement) for overall sample are reported.

Strong Evidence	Moderate Evidence	Minimal Evidence
Appropriate reliability estimates to support score interpretation are reported and the quality of reliability is compelling (i.e., most estimates meet or exceed 0.80).	Appropriate reliability estimates to support score interpretation are reported and the quality of reliability for some is not compelling (i.e., some estimates are between 0.70 and 0.80).	Appropriate reliability estimates to support score interpretation are not reported or the quality of reliability for many is poor (i.e., estimates are below 0.70).

- Reliability estimates are reported by subgroups, such as grade/age, gender, English learner status, exceptionality status, major racial/ethnic categories, and socio-economic status, numbers of participants included in each subgroup, and language backgrounds (e.g., those who speak language varieties) and those with disabilities (e.g., speech or hearing).

Strong Evidence	Moderate Evidence	Minimal Evidence
Reliability estimates to support appropriate score interpretation are disaggregated by most student groups, and the quality of reliability is compelling (i.e., most reliability estimates meet or exceed 0.80 for each reported group).	Reliability estimates to support appropriate score interpretation are disaggregated by some student groups and/or the quality of reliability is reasonable (i.e., some estimates are between 0.70 and 0.80).	Reliability estimates to support appropriate score interpretation are disaggregated by few or no student groups or the quality of reliability is weak (i.e., estimates are below 0.70)
and	and	or
If reliability estimates are lower for any student groups, there are appropriate suggested adjustments to administration and/or score interpretation.	If reliability estimates are lower for any student groups, there are appropriate suggested adjustments to administration and/or score interpretation.	Even though reliability estimates are lower for any student groups, there is no appropriate suggested adjustments to administration and/or score interpretation.

- **Content Validity-** Information on the content within the screening instrument, including information on items (development and selection; developmental appropriateness considering age/grade, linguistic and cultural aspects).

Strong Evidence	Moderate Evidence	Minimal Evidence
There is thorough and compelling evidence that the assessment items/tasks are appropriate measures of key skills, that the items/tasks are appropriate for the grade level, and that the mode of administration of the item/tasks is appropriate for the skills that are being assessed.	There is partial or reasonable evidence that the assessment items/tasks are appropriate measures of key skills, that the items/tasks are appropriate for the grade level, and that the mode of administration of the item/tasks is appropriate for the skills that are being assessed.	There is limited or unclear evidence that the assessment items/tasks are appropriate measures of key skills, that the items/tasks are appropriate for the grade level, or that the mode of administration of the item/tasks is appropriate for the skills that are being assessed.

- **Construct Validity-** Information showing that screening instrument measures the intended constructs for all student groups (e.g., age/grade differentiation, group differentiation [demographic and exceptionality status]).

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear and compelling evidence that the internal structure of the assessment and convergent or discriminant relations support the proposed score interpretation for overall group and subgroups.	There is partial or reasonable evidence that the internal structure of the assessment and convergent or discriminant relations support the proposed score interpretation for overall group and subgroups.	There is limited or unclear evidence that the internal structure of the assessment and convergent or discriminant relations support the proposed score interpretation for overall group and subgroups.

- Criterion Validity- Concurrent and predictive correlations (overall and for subgroups, including by age/grade, prior education experience, English language proficiency level, specific language background, and exceptionality status, where appropriate).

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear and convincing evidence that the total/composite score and/or relevant subscores are related to appropriate and meaningful external variables. and Concurrent or predictive correlations are reported for the overall group and most of the subgroups and results are in line with theory and evidence. Correlations are strongly aligned with intended outcomes and assessment time points (e.g., time of the year or time span between assessments in a longitudinal relation).	There is some and acceptable evidence that the total/composite score and/or relevant subscores are related to appropriate and meaningful external variables. and Concurrent or predictive correlations are reported for the overall group and some subgroups and results are in line with theory and evidence. Correlations are reasonably aligned with intended outcomes and assessment time points (e.g., time of the year or time span between assessments in a longitudinal relation).	There is limited or unconvincing evidence that the total/composite score and/or relevant subscores are related to appropriate and meaningful external variables. or Concurrent or predictive correlations are reported for the overall group but not most subgroups or may have limited alignment with theory and evidence. Correlations are weakly aligned with intended outcomes and assessment time points (e.g., time of the year or time span between assessments in a longitudinal relation).

- Criterion Validity- Classification accuracy: Specificity and sensitivity for identifying students' reading difficulty status, reported by the above-mentioned subgroups and with reference to language background, English language proficiency levels, exceptionality status, and prior education.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear and compelling evidence of classification accuracy, showing that the instrument appropriately identifies students who are and who are not at risk for reading difficulties by subgroups. Estimates of sensitivity and specificity are strongly aligned with intended outcomes and assessment time points.	There is some acceptable evidence of classification accuracy, showing that the instrument identifies students who are and who are not at risk for reading difficulties by subgroups. Estimates of sensitivity and specificity are reasonably aligned with intended outcomes and assessment time points.	There is limited or unconvincing evidence of classification accuracy, showing that the instrument identifies students who are and who are not at risk for reading difficulties by subgroups. Estimates of sensitivity and specificity are weakly aligned with intended outcomes and assessment time points.

- Criterion Validity- Types of decision rules such as benchmark goals and/or risk levels and associated evidence, reported by the above-mentioned subgroups and with reference to language background, English language proficiency levels, exceptionality status, and prior education.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear and compelling evidence for the decision rules (e.g., benchmark goals or risk levels) that inform a student's potential risk status for reading difficulties, by student subgroups.	There is some acceptable evidence for the decision rules (e.g., benchmark goals or risk levels) that inform a student's potential risk status for reading difficulties, by student subgroups.	There is limited or unconvincing evidence for the decision rules (e.g., benchmark goals or risk levels) that inform a student's potential risk status for reading difficulties by, student subgroups.

- Information on the participants who participated in collection of the reliability and validity data, including numbers of participants, demographic characteristics (e.g., grade/age, gender, race/ethnicity, exceptionality status, English learner status, socio-economic status, and those who speak language varieties), and geographic regions (including urbanicity).

Strong Evidence	Moderate Evidence	Minimal Evidence
There is compelling evidence that data from samples of students who are representative of Maryland's student population have informed assessment design, as well as reliability and validity analyses.	There is acceptable evidence that data from samples of students who are representative of Maryland's student population have informed assessment design, as well as reliability and validity analyses.	There is limited evidence that data from samples of students who are representative of Maryland's student population have informed assessment design, as well as reliability and validity analyses.

E.-F. Communication and Resources

Guidance and resources (E)

The extent to which the screening instruments offer useful guidance, resources, and professional development for the administration, interpretation of data, and reporting of results for populations that represent the student demographics of Maryland

- Guidance is well developed for administration of the instrument in cases that may require special considerations.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear, well-grounded guidance provided for administration of the instrument to students with disabilities and students at different levels of proficiency in the language of the instrument to support fair and accurate outcomes, including guidance about implementing evidence-based accommodations and training for those administering the instrument.	There is some guidance provided for administration of the instrument to students with disabilities and students at different levels of proficiency in the language of the instrument to support fair and accurate outcomes, including guidance about implementing evidence-based accommodations and/or training for those administering the instrument.	There is limited guidance or training provided for administration of the instrument to students with disabilities and students at different levels of proficiency in the language of the instrument to support fair and accurate outcomes.

- Resources are available, including professional development, for teachers, school and district leaders, and parents/guardians.

Strong Evidence	Moderate Evidence	Minimal Evidence
There are high-quality and user-friendly resources regarding the use of the instrument and professional development that are readily available, ongoing, and accessible to teachers, school and district leaders, and parents/guardians.	There are acceptable quality and user-friendly resources regarding the use of the instrument and professional development that are readily available, ongoing, and accessible to teachers, school and district leaders, and parents/guardians.	There are limited/no resources or low-quality/not user-friendly resources regarding the use of the instrument and professional development that are readily available, ongoing, and accessible to teachers, school and district leaders, and parents/guardians.

- Safeguards to protect student privacy and confidentiality.

Strong Evidence	Moderate Evidence	Minimal Evidence
There are thorough and high-quality guidance and safeguards for protecting student privacy and confidentiality.	There are partial and acceptable guidance and safeguards for protecting student privacy and confidentiality.	There are limited or incomplete guidance and safeguards for protecting student privacy and confidentiality.

Family Communication, resources, and Safeguards (F)

The extent to which the screening instruments offer educators and families useful guidance for next steps, including potential instructional responses, based on students' performances.

- How information about performance and relevant context factors is reported and analyzed for potential needs and next steps.

Strong Evidence	Moderate Evidence	Minimal Evidence
There is clear information about student performance, including guidelines for next steps that are responsive to student performance and context, and that offer educationally well-grounded advice.	There is reasonable information about student performance, including guidelines for next steps that are somewhat responsive to student performance and context, and that are moderately clear or usable.	There is limited information about student performance; guidelines for next steps are not responsive to student performance or context and/or lack clarity and usability.

- Resources available in multiple languages, with consideration of the languages of parents/guardians.

Strong Evidence	Moderate Evidence	Minimal Evidence
There are clear, usable, and easy to access resources for educators and families in multiple languages.	There are acceptable resources for educators and families in multiple languages, but accessibility, usability, or availability in multiple languages are limited.	There are limited resources for educators and families, or accessibility and usability are unclear and/or information is not available in multiple languages.

Appendix B: Recruitment and Review Timeline

Activity	Meeting Date
Panel recruitment and application review	September 2025
Panel members convene to receive training on the review process, review elements and evaluation criteria, and rubric.	October 2025
Panel Meeting (Deliberations) Panel members convene to develop the final list of screening instruments and release the screener information for selected screeners.	November 2025
Panel Meeting (Deliberations) Panel members convene to develop the final list of screening instruments and release the screener information for selected screeners.	December 2025
MSDE releases final report detailing the review process, and summaries of the review for each submitted screener.	March 2026
Updated “at-risk identifier” for Renaissance, added Appendix D, and removed PreK from Required Measure per COMAR 13A.03.08 chart.	June 2026

Appendix C: Definitions

DEFINITIONS

Classification Accuracy- Sensitivity and specificity for identifying students at-risk and not at-risk of reading difficulties, reported overall and by subgroup. Estimates must be aligned with intended outcomes and time points.

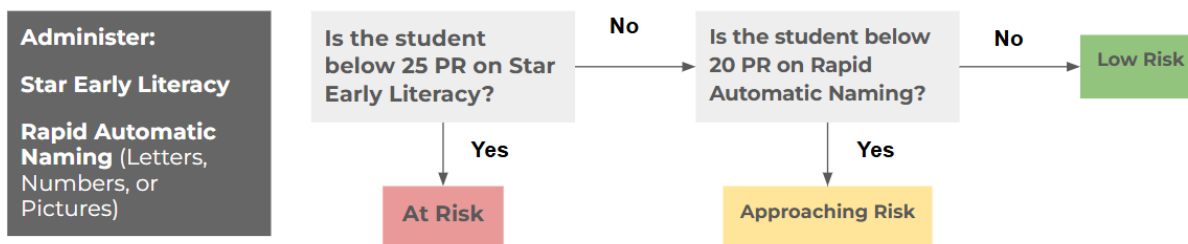
- **Content validity-** Information on the content of each task, including item development and selection; evidence that items are developmentally appropriate, linguistically and culturally responsive, and aligned with the State of Maryland's foundational reading skill expectations.
- **Construct validity-** Evidence that the instrument measures intended literacy constructs, supported by internal structure analyses and convergent/discriminant relationships. Validity must be demonstrated for the overall population and for Maryland's student subgroups.
- **Criterion validity-** Concurrent and predictive correlations with external measures of reading performance. Evidence should be reported overall and by subgroup (English proficiency level, disability status, prior educational experience). Correlations should align with intended outcomes and time points.
- **Decision Rules-** Evidence supporting benchmarks, cut-scores, and risk levels used to determine student risk, including justification for subgroup-specific considerations.
- **Dyslexia-** a specific learning disability characterized by difficulties in word reading and/or spelling that involve accuracy, speed, or both and vary depending on the orthography. These difficulties occur along a continuum of severity and persist even with instruction that is effective for the individual's peers. The causes of dyslexia are complex and involve combinations of genetic, neurobiological, and environmental influences that interact throughout development. Underlying difficulties with phonological and morphological processing are common but not universal, and early oral language weaknesses often foreshadow literacy challenges. Secondary consequences include reading comprehension problems and reduced reading and writing experience that can impede growth in language, knowledge, written expression, and overall academic achievement. Psychological well-being and employment opportunities also may be affected. Although identification and targeted instruction are important at any age, language and literacy support before and during the early years of education is particularly effective.
- **Psychometrics-** the science of measuring knowledge, skills, abilities, and academic traits through the development, validation, and analysis of assessment instruments. It applies statistical models to ensure tests are valid, reliable, and fair, enabling accurate, data-driven decisions about student performance and learning outcomes.
- **Reading Difficulties-** refers to a student's challenges with decoding, language comprehension, or both, which affect their ability to achieve grade-level reading proficiency.
- **Reliability-** Internal consistency, test-retest, alternate form, and inter-rater agreement for the overall sample. Reliability estimates must also be reported by subgroup (grade/age,

gender, English learner status, disability status, race/ethnicity, socioeconomic status, and language background).

- **Universal Reading Screener**- is an assessment administered three times per year (beginning, middle, and end) to identify or predict students at-risk for poor reading outcomes. It is typically brief and assesses skills such as phonological and phonemic awareness, phonics, fluency, and vocabulary, as developmentally appropriate.

Appendix D: Renaissance Scoring

Kindergarten



1st-3rd Grades

