Page 1

# Draft Reading Difficulties Risk Screener Selection Panel (RDRSSP) Proposed Review Elements and Evaluation Criteria

The RDRSSP’s Governing Policy, sections I.D. and I.E., specifies that RDRSSP will advise the State Board of Education (SBE) on a review process with evaluation criteria for the SBE’s adoption pursuant to California *Education Code* (*EC*) Section 53008, subdivisions (b) and (g)(1), for the RDRSSP to use to evaluate and approve a list of screening instruments by December 31, 2024. Section I.E specifies that the RDRSSP will advise the SBE on its recommended review process under *EC* Section 53008, subdivisions (b) and (g)(1), for the SBE’s adoption at the SBE’s meeting scheduled to be held on May 8–9, 2024.

The following review elements are provided for consideration of the aforementioned recommended review process and evaluation criteria.

*EC* Section 53008(g)(1)(A)–(E) specifies that, to support the adoption of high-quality screening instruments that minimize the overidentification or under-identification of pupils' risk of reading difficulties and offer meaningful information for follow up, the extent to which a screening instrument addresses the following factors be considered in the adoption of evaluation criteria:

1. Use of direct measurement, supplemented by other pupil data, to determine if a pupil is at risk of a reading difficulty, including dyslexia.
2. Measurement of domains that may predict dyslexia and other reading disorders, including, but not limited to, measures of 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.
3. Evidence that the tool is normed and validated using a contemporary multicultural and multilanguage sample of pupils, with outcome data for pupils whose home language is a language other than English as well as those who are native English speakers.
4. Integration of relevant pupil demographic information, such as home language, English language fluency, and access to prekindergarten education, to more fully understand a pupil's performance.
5. Guidance and resources for educators regarding how to administer screening instruments, interpret results, explain results to families, including in pupils' primary languages, and determine further educational strategies, assessments, diagnostics, and interventions that should be considered and that are specific to each type of pupil result. Guidance and resources provided pursuant to this subparagraph shall be informed by the English Language Arts/English Language Development Framework for California Public Schools developed pursuant to Section 60207 and the California Dyslexia Guidelines developed pursuant to
6. Section 56335, as well as knowledge of effective interventions for the specific needs of individual pupils, and shall reflect a tiered interventions model aligned with the Multi-Tiered Systems of Support.

Within this context, the following review elements are expected to be used for evaluating screeners. In cases where assessments are offered in languages other than English, the details outlined below should be supplied for each language of assessment. If information pertaining to any of the points below is unavailable, a justification should be provided.

## Description of Assessment Battery

1. Constructs and rationale: In line with *EC* Section 53008(g)(1)(A) and *EC* Section 53008(g)(1)(B), constructs directly measured at each grade level should be listed and described. The Panel must consider the extent to which these constructs include the following: 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.

Theoretical frameworks and prior research and evidence for the constructs measured and their developmental appropriateness at each level should be provided. For instruments in languages other than English, the rationale and evidence for how constructs have been modified as appropriate to the language should also be provided.

1. In line with *EC* Section 53008(g)(1)(C) and (D), descriptions of tasks should include the following:
   1. Their intended appropriate use
   2. For each task at each grade level, a description of the administration (individual or group administration; identity and 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
   3. For instruments in languages other than English, how tasks have been constructed to appropriately reflect relevant language features
2. In line with *EC* Section 53008(g)(1)(C) and (D), information on interpretation and appropriate use of students’ performance across tasks should include the following:
   1. Standard interpretation, including types of decision rules such as benchmark goals and/or risk levels
   2. Other considerations (student and family language background, language of instruction, proficiency in the language of the screening instrument, prior educational experience) and how these are taken into account in decisions about screener use and interpretation

## Psychometrics

In line with *EC* Section 53008(g)(1)(C) and (D), the following information should be provided for each screener, where appropriate by grade level and screener language:

Participants: 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 region (including urbanicity)

Reliability

1. Appropriate reliability estimates for different types of tasks (e.g., internal consistency, test–retest, alternate form, interrater agreement)
2. Reliability 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)]

Validity

1. Content validity: Information on the content of each task, including information on items (development and selection; developmental appropriateness considering age/grade, linguistic and cultural aspects)
2. Construct validity (e.g., age/grade differentiation, group differentiation [e.g., demographic and exceptionality status])
3. Criterion validity
   * Concurrent and predictive correlations (overall and for subgroups, including by age/grade, prior education experience, English language proficiency level and specific language background, where appropriate)
   * Information about the minimum language proficiency level necessary for the instrument to yield valid information in each assessed language
   * Classification accuracy: Specificity and sensitivity for identifying students’ reading difficulty status, reported by the above-mentioned subgroups and with reference to language background, English proficiency level, and prior education
   * 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, and prior education

## Communication and Resources

In line with *EC* Section 53008(g)(1)(E), the following information should be provided:

1. How information about student performance and relevant context factors are reported and decoded by the screening instrument to identify student needs and next steps
2. User interfaces and data management systems for entering and viewing scores, as relevant to various users such as teachers, school and district leaders, and parents/guardians
3. Resources available, including professional development, for teachers, school and district leaders, and parents/guardians
   1. Guidance, resources, and professional development for educators regarding how to administer screening instruments, interpret results, explain results to families, and determine further educational strategies, assessments, diagnostics, and interventions that should be considered specific to each type of pupil result, including the method and mode for delivery of these resources or training
   2. For students with limited English proficiency or English learners, guidance about the minimum English language proficiency level required for valid assessment results in English and additional guidance about English proficiency level used to determine assessment in languages other than English
   3. Resources available in multiple languages, with consideration of the languages of parents/guardians
4. Safeguards to protect student privacy and confidentiality
5. Alignment with *English Language Arts/English Language Development Framework for California Public Schools: Kindergarten Through Grade Twelve*  (*ELA/ELD Framework*) and the California Dyslexia Guidelines.
6. Feedback from users on their prior experience using screening instruments and/or participating in or using training and resources

## Proposed RDRSSP Evaluation Criteria

The proposed review elements outlined above will be evaluated according to the following criteria:

1. 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)
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students representative of the California student population
3. The extent to which the mode of administration for the screening instruments are appropriate for the students being evaluated (by grade level and student need)
4. 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
5. The extent to which the screening instruments offer useful guidance, resources, and professional development, for the administration, interpretation or data, and reporting of results for populations that represent the student demographics of California
6. The extent to which the screening instruments offer educators and families useful guidance for next steps, including potential instructional responses, based on students’ performances
7. The extent to which the screening instruments align with California guidance in the *ELA/ELD Framework* and the California Dyslexia Guidelines

The evaluation process will produce a description, for each approved screener, of its features and valid uses to inform local educational agency decision-making and use.

## Definitions Addendum

*For the purpose of implementing the Rubric, the Panel is using the following definitions regarding psychometrics, see sources below.*

Classification accuracy – the degree to which the assignment of test takers to specific categories (i.e., having risk of reading difficulties or not) is accurate; the degree to which false positive and false negative classifications are avoided.

Reliability – the extent of consistency of test scores, including the relationship among items within an assessment or between different administrations of the assessment, and among raters. Types of reliability include:

1. internal consistency reliability, which measures consistency across items
2. test-retest reliability, which measures consistency at different time points
3. alternate or parallel form reliability, which measures consistency between different forms of a test
4. inter-rater reliability, which measures consistency across different scorers or raters

Validity – the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.

Construct validity – the degree to which a test assesses the underlying theoretical construct that is intended to measure. It is frequently evaluated examining the extent to which scores on the test correlates with tests that are theoretically related to the construct (convergent validity) and the extent to which scores on the test do not correlate with tests that are theoretically unrelated to the construct (discriminant validity).

Content validity – the degree to which the content of an assessment (i.e., items) represent or cover the full range of content or domain test is intended to measure. This is typically evaluated by experts in a particular domain.

Criterion validity – the degree to which test scores and criterion performance results are related.

Concurrent validity – a type of criterion validity that estimates the degree to which an assessment or test correlates with other reliable and valid assessments of the same construct or a theoretically related construct, measured at the same point in time.

Predictive validity – a type of criterion validity that estimates the degree to which an assessment or instrument predicts the same construct or a theoretically related construct that is measured at a later time.

## Sources:

1. Joint Committee on the Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education
2. Regional Educational Laboratory Reliability and Validity Handout (<https://ies.ed.gov/ncee/rel/regions/midwest/pdf/training-and-coaching/ReliabilityValidityHandout-508.pdf>)
3. 3) Institute of Education Sciences Guide to Using a Research-Based Process to Review and Select Early Literacy Assessments (
4. <https://ies.ed.gov/ncee/rel/regions/midwest/pdf/training-and-coaching/ResearchBasedProcess-508.pdf)>

California Department of Education, April 2024