# California Reading Difficulties Risk Screener Selection Panel Rubric

Staff Recommendation to the State Board of Education on May 8, 2024

***Introduction***

The purpose of this rubric is to assist in selecting high-quality screening instruments that accurately identify students at risk of reading difficulties, while minimizing the risk of overidentification or underidentification, as specified in California Education Code (EC) Section 53008, subdivisions (b) and (g)(1). The assessment battery should include developmentally appropriate constructs that are based on theory and empirical evidence and that consider age, developmental level, background, and language acquisition, with strong psychometric properties and evidence that reflect and consider the diverse student population in California.

The rubric is designed for use by the RDRSSP Panel members during the review and evaluation of applications. It aligns directly with the Review Elements and Evaluation Criteria and is organized into three sections: Description of Assessment Battery, Psychometrics, and Communication and Resources. Each section contains evaluation criteria, and each criterion is evaluated using three evidence levels: strong, moderate, and minimal evidence. *The six evaluation criteria and their associated evidence statements collectively, rather than each criterion alone, offer a comprehensive assessment of the appropriateness of the instrument.*  Developmental appropriateness of the instrument is noted throughout the Description of Assessment Battery (Evaluation Criteria a, b, & c) and Psychometrics (Evaluation Criteria d) sections. Distinguishing features of each evidence level are evaluated with particular attention to the quality of evidence provided, as well as the thoroughness of the supporting documentation.

To ensure thorough, consistent, and objective evaluation across all applications, *the* *Panel will present, discuss, and deliberate on the strengths and weaknesses of each instrument, including its overall appropriateness* *for accurately measuring risk of reading difficulties.*Additionally, consistent with EC Section 53008, subdivision (c)(2), the Panel plans to develop and publish a brief summary of information about each recommended and approved screening instrument to inform local decision makers about the properties and affordances of each.

1. **DESCRIPTION OF ASSESSMENT BATTERY (a, b, c)**
2. 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)
	* 1. 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. |

* + 1. 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. |

1. **DESCRIPTION OF ASSESSMENT BATTERY (a, b, c) (CONT.)**
	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)
		1. 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. |

1. **DESCRIPTION OF ASSESSMENT BATTERY (a, b, c) (CONT.)**
2. 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)
	* 1. 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*.  |

1. **DESCRIPTION OF ASSESSMENT BATTERY (a, b, c) (CONT.)**
	1. 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)
		1. 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)*,* andincluding 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). |

1. **DESCRIPTION OF ASSESSMENT BATTERY (a, b, c) (CONT.)**

b. 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)

* + 1. 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. |

**1. DESCRIPTION OF ASSESSMENT BATTERY (a, b, c) (CONT.)**

* 1. 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
		1. Information about the minimum language proficiency level necessary.

| **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*. |

1. **PSYCHOMETRICS (d)**
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the California student population

*Reliability*

* + 1. 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). |

1. **PSYCHOMETRICS (d) (CONT.)**
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the California student population

*Reliability*

* + 1. 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).*and*If reliability estimates are lower for any student groups, there are appropriate suggested adjustments to administration and/or score interpretation. | 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).*and*If reliability estimates are lower for any student groups, there are appropriate suggested adjustments to administration and/or score interpretation. | 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)*or*Even though reliability estimates are lower for any student groups, there is no appropriate suggested adjustments to administration and/or score interpretation. |

1. **PSYCHOMETRICS (d) (CONT.)**
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the California student population

*Content Validity*

1. 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*

1. 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. |

1. **PSYCHOMETRICS (d) (CONT.)**
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the California student population

*Criterion Validity*

* + 1. 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 overallgroupand *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). |

1. **PSYCHOMETRICS (d) (CONT.)**
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the California student population

*Criterion Validity*

* + 1. 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. |

* + 1. 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. |

1. **PSYCHOMETRICS (d) (CONT.)**
2. The extent to which the screening instruments have been shown to be reliable and valid for populations of students who are representative of the California student population

*Representative of California Students*

1. 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 California’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 California’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 California’s student population have informed assessment design, as well as reliability and validity analyses. |

1. **COMMUNICATION AND RESOURCES (e, f)**
2. 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 California
	* 1. 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.  |

* + 1. 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* resourcesregarding the use of the instrument and professional development that are readily available, ongoing, and accessible to teachers, school and district leaders, and parents/guardians. |

1. **COMMUNICATION AND RESOURCES (e, f)**

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 California

1. 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. |

1. **COMMUNICATION AND RESOURCES (e, f)**
2. The extent to which the screening instruments offer educators and families useful guidance for next steps, including potential instructional responses, based on students’ performances
	* 1. 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*. |

* + 1. 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. |

California Department of Education, May 2024