# Page 1

# California Reading Difficulties Risk Screener Selection Panel Rubric––DRAFT

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. 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, including the developmental appropriateness at each grade level, to support the inclusion of all constructs/content measured by the instrument. | There is a *reasonable* theoretical framework and some evidence, including the developmental appropriateness at each grade level, to support the inclusion of all constructs/content measured by the instrument. | The theoretical framework or evidence is *limited or unconvincing*, including the developmental appropriateness at each grade level, to support the inclusion of all constructs/content measured by the instrument. |

* + 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** |
| --- | --- | --- |
| There are *clear and detailed* content specifications for each grade level aligned with theoretical rationale and empirical evidence that supports the inclusion of the constructs measured by the instrument. | There are *reasonable* content specifications for each grade level aligned with theoretical rationale and empirical evidence that supports the inclusion of the constructs measured by the instrument. | There is *limited or unconvincing* content specifications for each grade level aligned with theoretical rationale and empirical evidence that supports the inclusion of the constructs measured by the instrument. |

* + 1. 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

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

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. Descriptions of tasks should include their intended and appropriate use

| **Strong Evidence** | **Moderate Evidence** | **Minimal Evidence** |
| --- | --- | --- |
| There is a *clear* description about intended appropriate use of tasks based on convincing evidence. | There is a *clear* or *reasonable* description about intended appropriate use of tasks, but the evidence base is unclear. | There is a *limited* or *no* description about intended appropriate use of tasks or information is not based on evidence. |

* + 1. 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

| **Strong Evidence** | **Moderate Evidence** | **Minimal Evidence** |
| --- | --- | --- |
| There is *thorough and detailed* guidance for appropriate administration, including guidelines for administration and the appropriate testing environment and guidance for handling testing irregularities.  There is *thorough and detailed* guidance for scoring the items/tasks and interpreting and reporting scores.  There is a *clear and thorough* description, *well-grounded in evidence*, about how to administer and score the instrument, considering students’ age and developmental level, linguistic backgrounds (e.g., those who speak English dialects and varieties), and any known disabilities (e.g., speech or hearing). | There is *some and reasonable* guidance for appropriate administration, including guidelines for administration and the appropriate testing environment and guidance for handling testing irregularities.  There is *some or partial* guidance for scoring the items/tasks and interpreting and reporting scores.  There is a *partial* description, *grounded in evidence*, about how to administer and score the instrument, considering students’ age and developmental level, linguistic backgrounds (e.g., those who speak English dialects and varieties), and any known disabilities (e.g., speech or hearing). | There is *unclear or limited guidance* for appropriate administration, including guidelines for administration and the appropriate testing environment and guidance for handling testing irregularities.  There is *unclear or limited* guidance for scoring the items/tasks and interpreting and reporting scores.  There is an *unclear or* *limited* description, with *limited* *evidence* about how to administer and score the instrument, considering students’ age and developmental level, linguistic backgrounds (e.g., those who speak English dialects and varieties), or any known disabilities (e.g., speech or hearing). |

* + 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. 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, evidence-based* guidance about the minimum language proficiency level necessary for administration of the instrument. | There is reasonableguidance about minimum language proficiency level but *lacks clarity* and/or has *limited* information about evidence. | There is *limited or little* guidance about the minimum language proficiency necessary for the instrument to yield valid information or guidance is *not based* on evidence. |

1. 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
   1. *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** |
| --- | --- | --- |
| All reliability estimates *are appropriate for the types of tasks and the quality of reliability is compelling* (i.e., most estimates meet or exceed 0.80). | All reliability estimates *are appropriate for the types of tasks but the quality of reliability for some are not compelling* (i.e., some estimates are between .70 and .80.) | Some reliability estimates *are not appropriate for the types of tasks* or the *quality of reliability for some are poor* (i.e., estimates are below 0.70.) |

* + 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 appropriate for different task types 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 appropriate for different task types are disaggregated by some student groups and/or the quality of reliability is reasonable (i.e., some estimates are between .70 and .80.).  and  If reliability estimates are lower for any student groups, there are appropriate suggested adjustments to administration and/or score interpretation. | Reliability estimates appropriate for different task types 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. *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. |

* 1. *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. *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 *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 *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 *partial,* *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 *some or limited 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. 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 outcome and assessment time points. | There is *some and 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 that informs a student’s potential risk status in reading difficulties by subgroups. | There is *some and acceptable* evidence that informs a student’s potential risk status in reading difficulties by subgroups. | There is *limited or unconvincing* evidence that informs a student’s potential risk status in reading difficulties by subgroups. |

* 1. *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 reliability and validity evidence is drawn from samples of students who are representative of California students and geographic regions. | There is *acceptable* evidence that reliability and validity evidence is drawn from samples of students who are representative of California students and geographic regions. | There is *limited* evidence that reliability and validity evidence is drawn from samples of students who are representative of California students and geographic regions. |

1. 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. 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 instruments, professional development, and customer service and support that are readily available and accessible to teachers, school and district leaders, and parents/guardians. | There are *acceptable quality and user-friendly* resources regarding instruments, professional development, and customer service and support that are readily available and accessible to teachers, school and district leaders, and parents/guardians. | There are *limited/no resources or low-quality resources* regarding instruments, professional development, and customer service and support that are readily available and accessible to teachers, school and district leaders, and parents/guardians. |

* + 1. 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

| **Strong Evidence** | **Moderate Evidence** | **Minimal Evidence** |
| --- | --- | --- |
| There is an *easy to access and user-friendly* user interface and data management system for entering and viewing scores, including the ability to assign various user roles. | There is an *accessible* user interface and data management system for entering and viewing scores, but some functionality is *limited or not user-friendly*. | There is auser interface and data management system for entering and viewing scores, but accessibility and the functionality are *limited* and *not user-friendly*. |

* + 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. 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 and actionable* information about student performance, including guidelines for *actionable* 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. |

* + 1. Feedback from users on their prior experience using screening instruments, and/or participating in or using training and resources

| **Strong Evidence** | **Moderate Evidence** | **Minimal Evidence** |
| --- | --- | --- |
| There is *thorough and convincing* documentation that describes users’ positive feedback about their prior experience using the screening instrument. | There is *some and reasonable* documentation that describes users’ positive feedback about their prior experience using the screening instrument. | There is *limited or unconvincing* documentation that describes users’ positive feedback about their prior experience using the screening instrument. |

California Department of Education, April 2024