Page 1

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

| **Evaluation Criterion** | **Review Element** | **Evidence Summary** | **Strong Evidence** | **Moderate Evidence** | **Minimal/No Evidence** | **Rater Comments** |
| --- | --- | --- | --- | --- | --- | --- |
| **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)** | ***Constructs/Content Measured*** | Theoretical frameworks and evidence, including their developmental appropriateness at each grade level, should be provided for the constructs/content in the screening instrument.  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. | There is a clear theoretical framework and compelling evidence, including the developmental appropriateness at each grade level, to support the inclusion of all constructs/content measured by the instrument.  There are clear and detailed content specifications (e.g., blueprint, item 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 a clear 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.  There are reasonable content specifications (e.g., blueprint, item specifications) for each grade level aligned with theoretical rationale and empirical evidence that supports the inclusion of the constructs measured by the instrument. | Theoretical framework is unclear or lacks evidence to support the inclusion of all constructs/content measured by the instrument.  There is limited and/or unconvincing content specifications (e.g., blueprint, item specifications) for each grade level aligned with theoretical rationale and empirical evidence that supports the inclusion of the constructs measured by the instrument. | (To be completed) |
| **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)** | ***Constructs/Content Measured*** | For languages other than English, the rationale and evidence for how constructs have been modified as appropriate to the language should also be provided. | There is a clear rationale and compelling evidence to support how the constructs/content have been modified for other languages. | There is a reasonable rationale and some evidence of how constructs/content have been modified for other languages. | There is minimal or no rationale and limited evidence about whether or how constructs or content have been modified for other languages. | (To be completed) |
| **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)** | ***Description of Assessment Tasks (including their mode of administration)*** | 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, scoring procedures, and types of scores and their interpretation | There is a clear and evidence-based description about intended appropriate use of tasks.  There is thorough and detailed guidance for appropriate administration, including protocols or guidelines for administration and the appropriate testing environment and guidance for handling testing irregularities before, during, or after administration, administration time, administration format (e.g., computer-based, paper-based), administration approach (e.g., one-on-one, small group), and scoring procedures.  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 instruments, 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 description about intended appropriate use of tasks.  There is some guidance for appropriate administration, including protocols or guidelines for administration and the appropriate testing environment and guidance for handling testing irregularities before, during, or after administration, administration time, administration format (e.g., computer-based, paper-based), administration approach (e.g., one-on-one, small group), and scoring procedures.  There is clear guidance for scoring the items/tasks and interpreting and reporting scores.  There is partial or unclear description, grounded in evidence about how to administer and score instruments, considering at least some of the following: 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 limited or little description about intended appropriate use of tasks.  There is unclear or limited guidance for appropriate administration, including protocols or guidelines for administration and the appropriate testing environment and guidance for handling testing irregularities before, during, or after administration, administration time, administration format (e.g., computer-based, paper-based), administration approach (e.g., one-on-one, small group), and scoring procedures.  There is unclear or unconvincing guidance for scoring the items/tasks and interpreting and reporting scores.  There is little or unconvincing description, with limited evidence about how to administer and score instruments, 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). | (To be completed) |
| **(continued)**  **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)** | ***(continued)***  ***Description of Assessment Tasks (including their mode of administration)*** | 1. In languages other than English, how tasks have been constructed to appropriately reflect language features | For languages other than English, the screening instrument provides compelling evidence that supports how the tasks are constructed to appropriately reflect language features. | For languages other than English, the screening instrument provides some evidence that supports how the tasks are constructed to appropriately reflect language features but does not adequately address all language features. | For languages other than English, the screening instrument provides limited or unconvincing evidence that supports how the tasks are constructed to appropriately reflect language features or the screening instrument cannot be used in languages other than English. | (To be completed) |
| **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** | ***Sufficient Language Proficiency*** | Information about the minimum language proficiency level necessary for the instrument to yield valid information in each assessed language should be provided. | There is clear evidence-based guidance about the minimum language proficiency level necessary for administration of the instrument. | There is guidance 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. | (To be completed) |
| **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*** | Appropriate reliability estimates for different types of tasks (e.g., internal consistency, test–retest, alternate form, interrater agreement)  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) | There is technical documentation that clearly provides multiple estimates of reliability that are appropriate for the purpose of the assessment (e.g., internal consistency, test-retest, alternate form, interrater agreement), including reliability estimates that are disaggregated by student groups, such as grade, gender, race/ethnicity, exceptionality status, and language backgrounds [e.g., those who speak language varieties and those with disabilities (e.g., speech or hearing).  Most or all reliability estimates meet or exceed 0.70. Where reliability estimates are lower for any student groups, there are appropriate suggested adjustments to administration and/or score interpretation. | There is technical documentation that provides reliability estimates that are appropriate for the purpose of the assessment (e.g., internal consistency, test-retest, alternate form, interrater agreement), including reliability estimates that are disaggregated by student groups, such as grade, gender, race/ethnicity, exceptionality status, and language backgrounds [e.g., those who speak language varieties and those with disabilities (e.g., speech or hearing).  Most or all reliability estimates meet or exceed 0.60. Where reliability estimates are lower for any student groups, there is appropriate suggested adjustments to administration and/or score interpretation. | There is technical documentation that provides some reliability estimates (e.g., internal consistency, test-retest, alternate form, interrater agreement), and/or does not provide estimates that are disaggregated by student groups, such as grade, gender, race/ethnicity, exceptionality status, and language backgrounds [e.g., those who speak language varieties and those with disabilities (e.g., speech or hearing).  All or some reliability estimates fall below 0.60 without appropriate suggested adjustments to administration and/or score interpretation when reliability estimates are lower for some student groups. | (To be completed) |
| **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*** | Information on the content within the screening instrument, including information on items (development and selection; developmental appropriateness considering age/grade, linguistic and cultural aspects) | 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 some or partial 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 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. | (To be completed) |
| **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** | ***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]) | There is clear and compelling evidence that the internal structure of the assessment (e.g., item statistics, relation among test components) and convergent or discriminant relations support the proposed score interpretation for overall group and subgroups (age/grade differentiation and group differentiation). | There is some or partial evidence that the internal structure of the assessment (e.g., item statistics, relation among test components) and convergent or discriminant relations support the proposed score interpretation for overall group and subgroups (age/grade differentiation and group differentiation). | There is limited or unclear evidence that the internal structure of the assessment (e.g., item statistics, relation among test components) and convergent or discriminant relations support the proposed score interpretation for overall group and subgroups (age/grade differentiation and group differentiation). | (To be completed) |
| **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*** | 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) | 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. | There is some or partial evidence that the total/composite score and/or relevant subscores are related to appropriate and meaningful external variables.  And/Or  Concurrent or predictive correlations are reported for some subgroups and results are in line with theory and evidence. | There is limited or unclear evidence that the total/composite score and/or relevant subscores are related to appropriate and meaningful external variables.  And/Or  Concurrent or predictive correlations are reported for limited subgroups and/or may have limited alignment with theory and evidence. | (To be completed) |
| **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** | ***Classification Accuracy*** | Specificity and sensitivity for identifying students’ reading difficulty status  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 | There is clear and compelling evidence that informs a student’s potential risk status in reading difficulties.  There is clear and compelling evidence of classification accuracy analysis, 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 meet or exceed 0.80. | There is some evidence that informs a student’s potential risk status in reading difficulties.  There is evidence of classification accuracy analysis, showing that the instrument identifies students who are and who are not at risk for reading difficulties by subgroups. Estimates of sensitivity and specificity meet or exceed 0.70. | There is limited evidence that informs a student’s potential risk status in reading difficulties.  There is some or limited evidence of classification accuracy analysis, showing that the instrument identifies students who are and who are not at risk for reading difficulties by subgroups. Estimates of sensitivity and specificity fall below 0.70. | (To be completed) |
| **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*** | 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). | There is compelling evidence that reliability and validity evidence is drawn from samples of students who are representative of California students (race/ethnicity, gender, socioeconomic status, exceptionality status, students with varying English language proficiency levels, and those who speak language varieties), and geographic location. | There is some evidence that reliability and validity evidence is drawn from samples of students who are representative of California students (race/ethnicity, gender, socioeconomic status, exceptionality status, students with varying English language proficiency levels, and those who speak language varieties), and geographic location. | There is limited evidence that reliability and validity evidence is drawn from samples of students who are representative of California students (race/ethnicity, gender, socioeconomic status, exceptionality status, students with varying English language proficiency levels, and those who speak language varieties), and geographic location. | (To be completed) |
| **The extent to which the screening instruments offer useful guidance for administration and data interpretation and reporting** | ***Administration, Interpretation, and Reporting*** | Resources are available, including professional development, for teachers, school and district leaders, and parents/guardians  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  Safeguards to protect student privacy and confidentiality. | There are thorough and clear descriptions about available 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 is thorough and detailed guidance and safeguards for protecting student privacy and confidentiality. | There are some 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 is some guidance and safeguards for protecting student privacy and confidentiality. | There are limited or no 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, including easily accessible customer service and support.  There is limited or incomplete guidance and safeguards for protecting student privacy and confidentiality. | (To be completed) |
| **The extent to which the screening instruments offer educators and families useful guidance for next steps based on students’ performances** | ***Student Outcomes and Interventions*** | How information about performance and relevant context factors is reported and analyzed for potential needs and next steps  Resources available in multiple languages, with consideration of the languages of parents/guardians  Feedback from users on their experience using screening instruments, and/or participating in or using training and resources | There is clear and actionable information about student performance, including guidelines for actionable next steps that are responsive to student performance and context (instructional experience, language background, etc.) and that offer educationally well-grounded advice.  There are rich and easy to access and usable resources for educators and families in multiple languages.  There are sufficient opportunities for users to provide feedback about their experience using the screening instrument. | There is information about student performance, including guidelines for potential next steps that are somewhat responsive to student performance and context and moderately clear or usable.  There are resources for educators and families in multiple languages, but accessibility, usability, or availability in multiple languages are unclear.  There are some opportunities for users to provide feedback about their experience using the screening instrument. | 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.  There are resources for educators and families, but accessibility and usability are unclear and/or information is not available in multiple languages.  There are minimal opportunities for users to provide feedback about their experience using the screening instrument. | (To be completed) |

California Department of Education, March 2024 (Note: paper size is 8.5 by 14)