|  |
| --- |
| National Center on Intensive Intervention |
| Standard ProtocolFor Evaluating Behavior Screening Tools |

**NCII Screening Definition**

The National Center on Intensive Intervention defines screening as a process using tools with convincing evidence of classification accuracy, reliability, and validity to identify students who may require intensive intervention efforts to meet their academic, social, emotional, and/or behavioral needs.

## **Please Read and Complete Before You Start**

In order to know if your tool qualifies for review, please answer the following four questions regarding the research evidence for your tool:

|  |  |  |
| --- | --- | --- |
| 1. Can you provide direct evidence\* on the effects of using your tool?

*\*Direct evidence refers to data from a study that has been conducted based on the tool submitted for evaluation. Studies that use data from the use of another tool, even if it is similar, are considered indirect evidence and will not be considered as adequate evidence for the purposes of this review.*  | [ ]  Yes | [ ]  No |
| 1. Do you have the following classification data (either *a* or *b*) for your tool?
 |  |  |
| * 1. Specificity, sensitivity, positive predictive power, and negative predictive power

OR | [ ]  Yes | [ ]  No |
| * 1. Area Under the Curve (AUC) derived from a Receiver Operating Characteristic (ROC) curve analysis
 | [ ]  Yes | [ ]  No |
| 1. Are your classification data analyses conducted using cut points identifying students in need of intensive intervention (e.g., low risk contrasted with the highest level of risk on the measure)?
 | [ ]  Yes | [ ]  No |
| 1. Do the scores from your measure focus on a social, emotional, or behavioral need?
 | [ ]  Yes | [ ]  No |
| 1. Are there at least three months between the screening and your outcome measure for classification accuracy and predictive validity analysis?
 | [ ]  Yes | [ ]  No |

If you cannot answer YES to all of these questions, we will not review your tool. If you are able to answer YES to all of the above questions, then your tool qualifies for review. Please proceed to the following page and begin filling out the protocol.

**Please note:** If your tool assesses more than one domain of behavior, you must submit separate protocols for each sub-test that is used to identify students.

NCII staff will review all of your submitted materials to ensure that they adhere to the qualifications for review, stated above. If it is found that your submission packet needs substantial amount of supplemental information or is missing critical information, the entire packet will be returned to you. A revised protocol packet with additional information may be re-submitted.

Results of the review will be posted on the NCII website, in the Behavior Screening Tools Chart. **Once the review has begun, withdrawal of tools from the process will not be permitted**.

NCII staff are available to answer questions or to assist you completing the protocol for submission. Please contact the National Center on Intensive Intervention:

National Center on Intensive Intervention

American Institutes for Research

1000 Thomas Jefferson Street, NW

Washington, DC 20007

Website: [www.intensiveintervention.org](http://www.intensiveintervention.org)

E-mail: ToolsChartHelp@air.org

 NCII@air.org

## Marketing Language Agreement

In order to be eligible for review, you must read and sign this marketing language agreement.

By signing this agreement, I have indicated my understanding of the intent and purpose of the NCII tools chart, and my agreement to use language that is consistent with this purpose in any marketing materials that will be used to publicize my product’s presence and ratings on the chart.

Specifically, I understand the following:

1. The Technical Review Committee (TRC) rated each submitted tool against established criteria but did not compare it to other tools on the chart. The presence of a particular tool on the chart does not constitute endorsement and should not be viewed as a recommendation from either the TRC or the National Center on Intensive Intervention.
2. All tools submitted for review are posted on the chart, regardless of results. The chart represents all tools that were reviewed, not those that were “approved.”

When marketing my product, I will not use any language that is inconsistent with the above. Examples of inappropriate marketing language include, but may not be limited to, the following:

* 1. Reference to a “top-ranked” product in comparison to other products on the chart
	2. Reference to “approval” or “endorsement” of the product by the NCII

If the Center becomes aware of any marketing material on my product that violates this agreement, I understand that I risk removal of the product from the chart. I also understand that I may draft language and submit to Center staff for review in advance of releasing it, in order to ensure compliance with this agreement.

**I have read and understand the terms and conditions of this Agreement. By signing below, I signify my agreement to comply with all requirements contained herein.**

Signature Date

Print Name

Organization

## Section I: Basic Information

### A. Tool Information

1. Screening Tool Name:
2. Developer:
3. Publisher:
4. Publication Date:
5. Submission Contacts
	1. Primary Contact:

Title/Organization:

Email address:

Telephone:

* 1. Alternate Contact:

Title/Organization:

Email address:

Telephone:

### B. Descriptive Information

1. Provide brief (1-2 paragraph) overview and description of the tool.

1. What grade(s) does the tool target, if applicable? Check all that apply.

[ ]  Pre-K

[ ]  Kindergarten

[ ]  1st grade

[ ]  2nd grade

[ ]  3rd grade

[ ]  4th grade

[ ]  5th grade

[ ]  6th grade

[ ]  7th grade

[ ]  8th grade

[ ]  9th grade

[ ]  10th grade

[ ]  11th grade

[ ]  12th grade +

1. What age(s) does the tool target, if applicable? Check all that apply.

[ ]  0-4 years old

[ ]  5 years old

[ ]  6 years old

[ ]  7 years old

[ ]  8 years old

[ ]  9 years old

[ ]  10 years old

[ ]  11 years old

[ ]  12 years old

[ ]  13 years old

[ ]  14 years old

[ ]  15 years old

[ ]  16 years old

[ ]  17 years old

[ ]  18+ years old

1. What is the language of administration?

1. What broad domain/construct does the tool screen? (List all that apply)

### C. Acquisition Information

1. Where can your tool be obtained?

Website:

Address:

Phone number:

Email address:

1. Describe basic pricing plan and/or structure of the tools, including, as applicable: cost per student per year, start-up or other one-time costs, reoccurring costs, training cost, and what is included in each expense.

1. Provide information on what is included in the published tools, including information about special accommodations for students with disabilities.

## Section II: Usability

### A. Time, Administration, and Frequency

1. Who can be the rater/scorer? (Check all that apply.)

[ ]  General education teacher [ ]  Special education teacher

[ ]  Parent [ ]  Child [ ]  External observer

[ ]  Other school personnel (please specify):

[ ]  Other:

1. What is the administration setting? (Check all that apply.)

[ ]  General education classroom [ ]  Special education classroom

[ ]  School office [ ]  Recess [ ]  Lunchroom

[ ]  Home

[ ]  Other (please specify):

1. What is the assessment format? (Check all that apply.)

[ ]  Direct observation [ ]  Rating scale

[ ]  Checklist [ ]  Performance measure

[ ]  Other (please specify):

1. What is the administration context? (Check all that apply.)

[ ]  Large group [ ]  Small group

[ ]  Individual

[ ]  Other (please specify):

1. Can students be rated concurrently by one administrator?

[ ]  Yes (how many):\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ]  No

1. How long does it take to administer?

Minutes per student:

Minutes per total group:

1. How long does it take to score?

Minutes per student:

Minutes per total group:

Scoring is automatic:

1. Does your tool provide discontinue rules?

[ ]  Basals [ ]  Ceilings

[ ]  Other (please specify):

[ ]  Not provided

1. If relevant, how many alternate forms are available?

Number of alternate forms:

Per unit (grade/level/etc.):

1. Are norms available?

[ ]  Yes

[ ]  No

1. Are benchmarks available?

[ ]  Yes How many benchmarks per year?\_\_\_\_\_\_\_\_\_\_

 During which months? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ]  No

### B. Training

1. How long is tester training?

[ ]  Less than 1 hour of training

[ ]  1-4 hours of training

[ ]  4-8 hours of training

[ ]  Training not required

[ ]  Information not available

1. Are there minimum qualification of the examiner?

[ ]  Yes (please specify): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ]  No

1. Are training manuals and materials available?

[ ]  Yes [ ]  No

1. Are training manuals/materials field-tested?

[ ]  Yes [ ]  No

1. Are training manuals/materials included in cost of tools

[ ]  Yes [ ]  No (please describe training costs):\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Is there ongoing technical support available?

[ ]  Yes (please describe):\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ]  No

### C. Scoring

1. What types of scores are available? (Check all that apply.)

[ ]  Raw score [ ]  Standard score

[ ]  Percentile score [ ]  Grade equivalents

[ ]  IRT-based score [ ]  Age equivalents

[ ]  Stanines [ ]  Normal curve equivalents

[ ]  Developmental benchmarks [ ]  Developmental cut points

[ ]  Equated [ ]  Probability

[ ]  Lexile score [ ]  Error analysis

[ ]  Composite scores [ ]  Subscale/subtest scores

[ ]  Other (please specify):

1. Is a scoring key available?

[ ]  Yes (please describe): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[ ]  No

1. What is the basis for calculating standard & percentile scores?

[ ]  Age norms [ ]  Grade norms

[ ]  Stanine [ ]  Normal curve equivalents

1. Specify how raw scores are calculated and what comprises cluster/composite score.

1. Describe the tool’s approach to screening, behavior samples, test format, and/or scoring practices, including steps taken to ensure that it is appropriate for use with culturally and linguistically diverse populations and students with disabilities.

## Section III: Technical Standards

In the following sections please provided evidence supporting the technical adequacy of your tool. Submissions must report data separately for each span of grade levels that is targeted by the screening instrument, in accordance with developer guidelines about target grade spans or ranges (e.g., K-1 or K-3). Data must also be reported separately by informant, if appropriate for the tool (e.g., teacher, parent, student). Evidence will be rated and reported on the chart separately for each possible combination of grade span and informant (e.g., K-1 teacher, K-1 parent). In cases where data are not available for one or more grades that fall within the grade span targeted by the tool, or one of the available informant forms, the TRC will give a rating of “—“ to indicate “data not available.”

### Technical Standard 1. Classification Accuracy

1. Please provide the following classification data for your tool disaggregated by informant, grade span, time of year, outcome measure.

Please note, we will only rate information on two different criterion measures. If you provide information from more than two criterion measures, we will report but not rate that additional information.

The cut points for this analysis should be aligned with students needing intensive intervention. If you provide information on cut points not associated with students needing intensive intervention, we will report but not rate that additional information.

Copy additional sections as needed.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Students Actually“At-Risk” | Students Actually“Not At-Risk” | Total |
| Students Classified as “At-Risk” | True Positive | False Positive |  |
|  | a | b | a + b |
| Students Classified as “Not At-Risk” | False Negative | True Negative |  |
|  | c | d | c + d |
|  |  |  | N = a+b+c+d |

1. Grade Level/Span:
2. Informant:
3. Time of Year:
4. Outcome:
5. What were the cut-points?
6. What is the base rate in the sample for children requiring intensive intervention?

1. What is the base rate in the sample for children considered at-risk, including those with the most intensive needs?
2. False Positive Rate [b/(b+d)]:
3. False Negative Rate [c/(a+c)]:
4. Sensitivity [a/(a+c)]:
5. Specificity [d/(b+d)]:
6. Positive Predictive Power [a/(a+b)]:
7. Negative Predictive Power [d/(c+d)]:
8. Overall Classification Rate [(a+d)/(a+b+c+d)]:
9. Area Under the Curve (AUC):
10. Please provide specificity for the following three sensitivity levels:
	1. Sensitivity 90%, Specificity:
	2. Sensitivity 80%, Specificity:
	3. Sensitivity 70%, Specificity:
11. Do you provide classification data that are disaggregated by gender, race, or ethnicity? If so, complete below for each subgroup group for which you provide disaggregated classification data. Copy additional pages as needed.

Subgroup: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |  |
| --- | --- | --- | --- |
|  | Students Actually“At-Risk” | Students Actually“Not At-Risk” | Total |
| Students Classified as “At-Risk” | True Positive | False Positive |  |
|  | a | b | a + b |
| Students Classified as “Not At-Risk” | False Negative | True Negative |  |
|  | c | d | c + d |
|  |  |  | N = a+b+c+d |

1. Grade Level:
2. Informant:
3. Time of Year:
4. Outcome:
5. What were the cut-points?
6. What is the base rate in the sample for children requiring intensive intervention?

1. What is the base rate in the sample for children considered at-risk, including those with the most intensive needs?
2. False Positive Rate [b/(b+d)]:
3. False Negative Rate [c/(a+c)]:
4. Sensitivity [a/(a+c)]:
5. Specificity [d/(b+d)]:
6. Positive Predictive Power [a/(a+b)]:
7. Negative Predictive Power [d/(c+d)]:
8. Overall Classification Rate [(a+d)/(a+b+c+d)]:
9. Area Under the Curve (AUC):
10. Please provide specificity for the following three sensitivity levels:
	1. Sensitivity 90%, Specificity:
	2. Sensitivity 80%, Specificity:
	3. Sensitivity 70%, Specificity:
11. Please provide a description of the study sample on which the preceding data are based.

National representation:

Northeast: [ ]  New England [ ]  Middle Atlantic

Midwest: [ ]  East North Central [ ]  West North Central

South: [ ]  South Atlantic [ ]  East South Central [ ]  West South Central

West: [ ]  Mountain [ ]  Pacific

Local representation (please describe, including number of states):

Date:

Size:

Gender (Percent): Male: \_\_\_\_\_ Female: \_\_\_\_\_ Unknown: \_\_\_\_\_

SES (Percent, measured as free or reduced-price lunch): \_\_\_\_\_

Eligible for free or reduced-price lunch: \_\_\_\_\_

Other SES Indicators: \_\_\_\_\_

Race/Ethnicity (Percent):

White, Non-Hispanic: \_\_\_\_\_

Black, Non-Hispanic: \_\_\_\_\_

Hispanic: \_\_\_\_\_

American Indian/Alaska Native: \_\_\_\_\_

Asian/Pacific Islander: \_\_\_\_\_

Other: \_\_\_\_\_

Unknown: \_\_\_\_\_

Disability classification (Please describe): \_\_\_\_\_

First language (Please describe): \_\_\_\_\_

Language proficiency status (Please describe): \_\_\_\_\_

1. Describe the criterion (outcome) measure(s) including the degree to which it is independent from the screening measure.

1. Describe how the classification analyses were performed and cut-points determined. Describe how the cut points align with (a) at-risk and/or (b) intensive need.

1. Were the children in the study involved in an intervention in addition to typical classroom instruction between the screening measure and outcome assessment?

□ Yes □ No

If yes, please describe the intervention, what children are receiving the intervention, and how they were chosen.

### Technical Standard 2. Reliability

In the section below, describe the reliability analyses conducted, and provide results. You may report more than one type of reliability (e.g., model-based, internal consistency, inter-rater reliability); however, you must also justify the appropriateness of the method used given the type and purpose of the tool.

Please ensure that you submit evidence for each informant (e.g. rater/ observer) and individual grade level targeted by the tool. If you fail to submit data for a targeted grade level, that grade will receive a “dash” rating for this GOM

1. Offer a justification for each type of reliability reported, given the type and purpose of the tool.

1. Describe the sample(s), including size and characteristics, for each reliability analysis conducted.

1. Describe the analysis procedures for each reported type of reliability.

1. In the chart below, report the reliability of performance level score (e.g., model-based, internal consistency, inter-rater reliability).

| Type of Reliability | Informant | Age or Grade | n | Coefficient | Confidence Interval |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Does your manual cites other published reliability studies?

[ ]  Yes [ ]  No

If yes, provide citations for additional published studies.

1. Do you provide, in your user’s manual, reliability data that are disaggregated by diverse populations (e.g., race-ethnicity, ELL students)? If so, complete below for each group for which you have disaggregated reliability data.

If you include inter-rater reliability, provide formula.

| Type of Reliability | Informant | Subgroup | Age or Grade | n | Coefficient | Confidence Interval |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Does your manual cites other published disaggregated reliability studies?

[ ]  Yes [ ]  No

If yes, provide citations for additional published studies.

Copy additional pages as needed.

### Technical Standard 3. Validity

In the section below, describe the validity analyses conducted, and provide results. You may report more than one type of validity (e.g., concurrent, predictive, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, and/or evidence based on consequences of testing), and more than one criterion measure. However, you must justify the choice of analysis and criterion measures given the theoretical assumptions about the relationship between your tool and other, similar constructs.

Please ensure that you submit evidence for each informant (e.g. rater/scorer) and individual grade level targeted by the tool. If you fail to submit data for a targeted grade level, that grade will receive a “dash” rating for this GOM.

1. Describe each criterion measure used and explain why each measure is appropriate, given the type and purpose of the tool. (NOTE: To support validity and generalizability, the TRC prefers and strongly encourages criterion measures that are ***external to the progress monitoring system***. If internal measures are used, please include a description of what provisions have been taken to address the limitations of this method, such as possible method variance or overlap of item samples.).

1. Describe the sample(s), including size and characteristics, for each validity analysis conducted.

1. Describe the analysis procedures for each reported type of validity.

1. In the chart below, report validity information for the performance level score (e.g., concurrent, predictive, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, and/or evidence based on consequences of testing), and the criterion measures.

| Type of Validity | Informant | Age or Grade | Test or Criterion | n | Coefficient | Confidence Interval |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Results for other forms of validity (e.g. factor analysis) not conducive to the table format.

Does your manual cites other published validity studies?

[ ]  Yes [ ]  No

If yes, provide citations for additional published studies.

1. Describe the degree to which the provided data support the validity of the tool.

1. Do you provide, in your user’s manual, validity data that are disaggregated by diverse populations (e.g., race-ethnicity, ELL students)? If so, complete below for each group for which you have disaggregated validity data.

| Type of Validity | Informant | Subgroup | Age or Grade | Test or Criterion | n | Coefficient | Confidence Interval |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Results for other forms of disaggregated validity (e.g. factor analysis) not conducive to the table format.

Does your manual cites other published disaggregated validity studies?

[ ]  Yes [ ]  No

If yes, provide citations for additional published studies.

Copy additional pages as needed.

### Technical Standard 4. Cross-Validation

1. Has a cross-validation study been conducted? *In a cross-validation study, previously identified classification rules are used to predict risk and non-risk in the new sample.*

[ ]  Yes [ ]  No

The cut points for this analysis should be aligned with students needing intensive intervention. If you provide information on cut points not associated with students needing intensive intervention, we will report but not rate that additional information.

If yes, provide the following classification data for the cross validation study.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Students Actually“At-Risk” | Students Actually“Not At-Risk” | Total |
| Students Classified as “At-Risk” | True Positive | False Positive |  |
|  | a | b | a + b |
| Students Classified as “Not At-Risk” | False Negative | True Negative |  |
|  | c | d | c + d |
|  |  |  | N = a+b+c+d |

1. Grade Level:
2. Informant:
3. Time of Year:
4. Outcome:
5. What were the cut-points?
6. What is the base rate in the sample for children requiring intensive intervention?

1. What is the base rate in the sample for children considered at-risk, including those with the most intensive needs?
2. False Positive Rate [b/(b+d)]:
3. False Negative Rate [c/(a+c)]:
4. Sensitivity [a/(a+c)]:
5. Specificity [d/(b+d)]:
6. Positive Predictive Power [a/(a+b)]:
7. Negative Predictive Power [d/(c+d)]:
8. Overall Classification Rate [(a+d)/(a+b+c+d)]:
9. Area Under the Curve (AUC):
10. Please provide specificity for the following three sensitivity levels:
	1. Sensitivity 90%, Specificity:
	2. Sensitivity 80%, Specificity:
	3. Sensitivity 70%, Specificity:
11. Please provide a description of the cross-validation study sample on which the preceding data are based.

National representation:

Northeast: [ ]  New England [ ]  Middle Atlantic

Midwest: [ ]  East North Central [ ]  West North Central

South: [ ]  South Atlantic [ ]  East South Central [ ]  West South Central

West: [ ]  Mountain [ ]  Pacific

Local representation (please describe, including number of states):

Date:

Size:

Gender (Percent): Male: \_\_\_\_\_ Female: \_\_\_\_\_ Unknown: \_\_\_\_\_

SES (Percent, measured as free or reduced-price lunch): \_\_\_\_\_

Eligible for free or reduced-price lunch: \_\_\_\_\_

Other SES Indicators: \_\_\_\_\_

Race/Ethnicity (Percent):

White, Non-Hispanic: \_\_\_\_\_

Black, Non-Hispanic: \_\_\_\_\_

Hispanic: \_\_\_\_\_

American Indian/Alaska Native: \_\_\_\_\_

Asian/Pacific Islander: \_\_\_\_\_

Other: \_\_\_\_\_

Unknown: \_\_\_\_\_

Disability classification (Please describe): \_\_\_\_\_

First language (Please describe): \_\_\_\_\_

Language proficiency status (Please describe): \_\_\_\_\_

1. Describe the criterion (outcome) measure(s) including the degree to which it is independent from the screening measure.

1. Describe how the cross-validation analyses were performed and cut-points determined. Describe how the cut points align with (a) at-risk and/or (b) intensive need.

1. Were the children in the cross-validation study involved in an intervention in addition to typical classroom instruction between the screening measure and outcome assessment?

[ ]  Yes [ ]  No

If yes, please describe the intervention, what children are receiving the intervention, and how they were chosen.

### Technical Standard 5. Bias Analysis

1. Have you conducted additional analyses related to the extent to which your tool is or is not biased against subgroups (e.g., race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners)? Examples might include Differential Item Functioning (DIF), or invariance testing in multiple-group confirmatory factor models.

[ ]  Yes [ ]  No

If yes, please provide the following information for each grade level/span and informant version of your assessment.

* 1. Describe the method used to determine the presence or absence of bias.

* 1. Describe the subgroups for which bias analyses were conducted.

* 1. Describe the results of the bias analyses conducted, including data and interpretative statements. Include magnitude of effect (if available) if bias is identified.