# Protocol for Evaluating Academic Progress Monitoring Tools

**National Center on Intensive Intervention**

October 2017

The National Center on Intensive Intervention defines progress monitoring as repeated measurement of student performance over the course of intervention to index/quantify responsiveness to intervention and to thus determine, on an ongoing basis, when adjustments to the program are needed to improve responsiveness. When the need for a program adjustment is determined, supplementary data sources (e.g., functional behavior assessments, diagnostic academic assessments, informal observations, work samples) or more fine-grained data available within the repeated measurement samples are used to decide the most productive strategies for altering intervention. The purpose of this progress monitoring is to design an individualized intervention that optimizes student outcomes.

**Please Read Before You Start**

***Q1. Are there minimum criteria that my tool must meet in order to qualify for review?***

* **Yes**. The TRC will only review submissions that meet the following five criteria:
1. Measure must target academic functioning.
2. Measure must involve formative assessment (i.e. repeated administration), with the intended purpose of progress monitoring.
3. Measure must include, but is not limited to, monitoring of individual student performance.
4. Evidence under consideration must be **direct** evidence; in other words, it is derived from data collected on the tool being submitted for review. Indirect evidence, or data collected on tools similar to the tool being reviewed, will not be accepted.
5. Evidence of reliability and validity must be provided for a grade level to be reviewed. Data spanning multiple grade levels is not accepted.

Center staff will review this submission upon receipt, to ensure that these minimum criteria are met. Only submissions that are determined to meet all five criteria will be assigned for review.

1. ***My tool does not have enough alternate forms to conduct progress monitoring at least weekly, as required for data-based individualization. Can I still submit my tool for review?***
* **Yes.** The TRC will review tools with fewer than 20 alternate forms but still highly recommends that at least 20 alternate forms be available when using progress monitoring tools for DBI purposes. The TRC will take this into account when making ratings, and vendors will be expected to justify that the number of alternate forms available is sufficient.
1. ***My progress monitoring tool assesses multiple domains of academic performance (e.g., reading vs. mathematics or mathematics computation vs. concepts or applications). Do I need a separate protocol for each domain?***
* **Yes**. The Center recognizes that for products designed to measure progress in multiple academic domains, some of the information to be submitted in the protocol will be the same. However, the tool for each academic domain or subcomponent within a domain will be evaluated and reported separately on our tools chart. Therefore, if your tool assesses more than one domain or subcomponent, you **MUST** submit **separate** protocols for **EACH** domain or subcomponent. For example, if your tool measures subcomponents of reading, such as letter name fluency, letter sound fluency, and passage reading fluency, you must submit a separate protocol for each.
1. ***The protocol requires information that is already included in a technical report or research study. Can I submit this study instead of filling out the protocol?***

**No**. Technical reports and relevant research papers may be submitted as supporting information, but you ***MUST COMPLETE THE FULL PROTOCOL***. Reviewers will use the information in the protocol to make their judgments. They are not expected to search for and find additional information in accompanying materials.

1. ***The protocol requires information that is not currently available. Can I still submit my progress monitoring tool?***
* **Yes**. The Protocol for Evaluating Progress Monitoring Tools is designed to collect comprehensive and detailed information on the submitted progress monitoring tools to ensure rigorous evaluation of tools. Therefore, tools that are undergoing improvements or are in an early phase of development may not have all the information the protocol asks for. Please provide as much information as is available and possible.

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 then be submitted.

1. ***Can I withdraw my tool from the review process?***
* **No.** Results of the review will be posted on the Center’s website, in the Progress Monitoring Tools Chart. ***Once the review has begun, withdrawal from the process is not permitted.***
1. ***I am not familiar with some of the terms in the protocol, and thus I am not sure what information I should provide. What should I do?***
* Center staff is available to answer your questions and to assist you in completing the protocol for submission. Please contact the National Center on Intensive Intervention at the address below:

**National Center on Intensive Intervention**

American Institutes for Research

1000 Thomas Jefferson Street NW

Washington, DC 20007

E-mail: ToolsChartHelp@air.org

## **Marketing Language Agreement**

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

By signing this agreement I have indicated my understanding of the intent and purpose of the NCII tools charts, 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 NCII 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 NCII 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

* 1. **Tool Information**
1. 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:

* 1. **Descriptive Information**
1. Description of 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 [ ]  10 years old [ ]  16 years old

[ ]  5 years old [ ]  11 years old [ ]  17 years old

[ ]  6 years old [ ]  12 years old [ ]  18+ years old

[ ]  7 years old [ ]  13 years old

[ ]  8 years old [ ]  14 years old

[ ]  9 years old [ ]  15 years old

1. The tool is intended for use with the following student populations (check all that apply):

[ ]  Students in general education [ ]  Students with disabilities [ ]  English language learners

1. What dimensions does tool assesses? Check all that apply.

|  |
| --- |
| **Reading** |
| [ ]  Global Indicator of Reading Comprehension[ ]  Listening Comprehension[ ]  Vocabular[ ]  Phonemic Awareness[ ]  Decoding[ ]  Passage Reading[ ]  Word Identification[ ]  Comprehension |

|  |
| --- |
| **Spelling & Written Expression** |
| [ ]  Global Indicator of Spelling Competence[ ]  Global Indicator of Written Expression Competence |

|  |
| --- |
| **Mathematics** |
| [ ]  Global Indicator of Mathematics Comprehension[ ]  Early Numeracy[ ]  Mathematics Concepts[ ]  Mathematics Computation[ ]  Mathematics Application[ ]  Fractions[ ]  Algebra |

|  |
| --- |
| **Other** |
| List specific domain, skills or subtests: |

* 1. **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: Development and Administration

1. **Time, Administration, and Frequency**
2. What is the assessment format? Check all that apply.

[ ]  Individual [ ]  Group [ ]  Computer-administered

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. How many alternate forms are available?

 [#] alternate forms per [grade/level/unit]

1. **Training**
2. How long is tester training?

[ ]  Less than 1 hour of training

[ ]  1-4 hours of training

[ ]  4-8 hours of training

[ ]  8 or more 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

1. **Scoring**
2. How is scoring conducted?

[ ]  By hand [ ]  Computer-scored

1. Do you provide basis for calculating performace level scores?

[ ]  Yes [ ]  No

1. What types of performance level 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. What is the basis for calculating performance level standard and percentile scores?

[ ]  Age norms [ ]  Grade norms [ ]  Stanines [ ]  Normal curve equivalent

1. What is the scoring structure? Specify ow raw scores are calculated and what the cluster/composite score comprises.

1. Do you provide basis for calculating slope (e.g., amount of improvement per unit in time)?

[ ]  Yes [ ]  No

1. Do you provide benchmarks for the slopes?

[ ]  Yes [ ]  No

1. Fo you provide percentile ranks for the slopes?

[ ]  Yes [ ]  No

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

[ ]  Age norms [ ]  Grade norms [ ]  Stanines [ ]  Normal curve equivalent

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

1. Describe the tool’s approach to progress monitoring, 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.

1. **Rates of Improvement and End of Year Benchmarks**

1. Is minimum acceptable growth (slope of improvement or average weekly increase in score by grade level) specified in your manual or published materials?

[ ]  Yes [ ]  No

If yes, specify the growth standards:

1. Are benchmarks for minimum acceptable end-of-year performance specified in your manual or published materials?

[ ]  Yes [ ]  No

If yes, specify the end-of-year performance standards:

1. What is the basis for specifying minimum acceptable growth and end of year benchmarks?

[ ]  Norm-referenced [ ]  Criterion-referenced [ ]  Other

If norm-referenced, describe the normative profile.

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: \_\_\_\_\_

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. Do you provide, in your user’s manual, norms which are disaggregated by race or ethnicity? If so, for which race/ethnicity? Check all that apply.

[ ]  White, Non-Hispanic [ ]  Black, Non-Hispanic

[ ]  Hispanic [ ]  American Indian/Alaska Native

[ ]  Asian/Pacific Islander [ ]  Other

[ ]  Unknown

1. If criterion-referenced, describe procedure for specifying criterion for adequate growth and benchmarks for end-of-year performance levels (attach documentation).

1. Describe any other procedures for specifying adequate growth and minimum acceptable end of year performance.

## Section III: Technical Information

1. **Foundational Psychometric Standards**

**A1. Reliability of Performance Level Score**

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. It is expected that the sample for these analyses represents the general student population (or intended population of the tool if it differs from the general population).

Please ensure that you submit evidence for each individual grade level targeted by the tool.

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** | **Age or Grade** | **n** | **Coefficient** | **Confidence Interval** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Manual cites other published reliability studies: [ ]  Yes [ ]  No

Provide citations for additional published studies.

1. Do you provide, in your user’s manual, reliability for the performance level score that is disaggregated by subgroups (e.g., race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners)? If so, complete below for each subgroup for which you provide disaggregated reliability for the performance level score data.

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

Manual cites other published reliability studies: [ ]  Yes [ ]  No

Provide citations for additional published studies.

**A2. Validity of Performance Level Score**

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. It is expected that the sample for these analyses represents the general student population (or intended population of the tool if it differs from the general population).

Please ensure that you submit evidence for each individual grade level targeted by the tool.

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** | **Age or Grade** | **Test or Criterion** | **n** | **Coefficient** | **Confidence Interval** |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

 Results for other forms of validity not conducive to the table format:

Manual cites other published reliability studies: [ ]  Yes [ ]  No

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 for the performance level score that is disaggregated by subgroups (e.g., race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners)? If so, complete below for each subgroup for which you provide disaggregated validity for the performance level score data.

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

 Results for other forms of validity not conducive to the table format:

Manual cites other published reliability studies: [ ]  Yes [ ]  No

Provide citations for additional published studies.

**A3. Bias Analyses**

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,

* 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 has been identified.

1. **Progress Monitoring with Intensive Population**

Note: In this section, it is expected that the sample for all analyses represent students in need of intensive intervention. When describing your sample, provide evidence that students demonstrate this need. Convincing evidence that children were in need of intensive intervention may include one or more of the following: all students below the 30th percentile on local or national norm or sample mean below 25th percentile on local or national test; students have an IEP with reading goals or math goals that are consistent with the tool; or students are non-responsive to Tier 2 instruction.

**B1. Sensitivity to Student Learning: Reliability of Slope**

In the section below, describe the reliability of the slope analyses conducted.

Please ensure that you submit evidence for each 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 standard.

1. Describe the sample, including size and characteristics:

1. Describe the frequency of measurement (for each student in the sample, report how often data were collected and over what span of time):

1. Describe the analysis procedures:

1. In the chart below, report reliability of the slope (e.g., ratio of true slope variance to total slope variance) by grade level (if relevant).

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

Manual cites other published reliability studies: [ ]  Yes [ ]  No

Provide citations for additional published studies.

1. Do you provide, in your user’s manual, reliability for the slope that is disaggregated by subgroups (e.g., race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners)? If so, complete below for each race/ethnicity for which you provide disaggregated reliability for the slope data.

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

Manual cites other published reliability studies: [ ]  Yes [ ]  No

Provide citations for additional published studies.

**B2. Sensitivity to Student Learning: Validity of Slope**

In the section below, describe the predictive validity for the slope of improvement (correlation between the slope and achievement outcome) analyses. You may report more than one criterion measure. However, you must justify the choice criterion measures given the theoretical assumptions about the relationship between your tool and other, similar constructs.

Please ensure that you submit evidence for each 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 standard.

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, including size and characteristics:

1. Describe the frequency of measurement (for each student in the sample, report how often data were collected and over what span of time):

1. Describe the analysis procedures:

1. In the chart below, report predictive validity information for the slope of improvement (correlation between the slope and achievement outcome).

\*\*Please note, the TRC suggests controlling for initial level when the correlation for slope without such control is not adequate.

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

Manual cites other published reliability studies: □ Yes □ No

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, predictive validity information for the slope of improvement that is disaggregated by subgroups (e.g., race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners)? If so, complete below for each race/ethnicity for which you provide disaggregated predictive validity for the slope of improvement data.

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

Manual cites other published reliability studies: [ ]  Yes [ ]  No

Provide citations for additional published studies.

**B3. Alternate Forms**

Provide evidence that alternate forms are of equal and controlled difficulty or, if IRT based, provide evidence of item or ability invariance (provide or attach documentation of direct evidence).

Please ensure that you submit evidence for each 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 standard.

1. Describe the sample for these analyses, including size and characteristics:

1. What is the number of alternate forms of equal and controlled difficulty?

1. If IRT based, provide evidence of item or ability invariance (attach documentation):

1. If computer administered, how many items are in the item bank for each grade level?

1. If your tool is computer administered, please note how the test forms are derived instead of providing alternate forms:

**B4. Decision Rules for Setting and Revising Goals**

1. Does your manual or published materials specify validated decision rules for how to set and revise goals?

Please ensure that you submit evidence for each 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 standard.

[ ]  Yes [ ]  No

If yes, specify the decision rules:

1. What is the evidentiary basis for these decision rules?

NOTE: The TRC expects evidence for this standard to include an empirical study that compares a treatment group to a control and evaluates if student outcomes increase when decision rules are in place.

**B5. Decision Rules for Changing Instruction**

1. Does your manual or published materials specify validated decision rules for when changes to instruction need to be made?

Please ensure that you submit evidence for each 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 standards.

[ ]  Yes [ ]  No

If yes, specify the decision rules:

1. What is the evidentiary basis for these decision rules?

NOTE: The TRC expects evidence for this standard to include an empirical study that compares a treatment group to a control and evaluates if student outcomes increase when decision rules are in place.