National Center on
INTENSIVE INTERVENTION
at American Institutes for Research ■

AIR®
AMERICAN INSTITUTES FOR RESEARCH®

IDEAs
that Work
U.S. Office of Special
Education Programs

Behavior Progress Monitoring Tools Chart Rating Rubric

*Please note that the following rubrics are applied separately for **each sub-scale, grade level/span, and informant** targeted by the progress monitoring tool.*

Tools Chart Tab 1: Performance Level Standards

Note: For all standards in Tab 1, it is expected that evidence is drawn from a sample that is representative of students across all performance levels. Evidence not meeting this criterion will receive a rating of "—" to indicate "data not available."

1. Reliability

Rating	Definition
Full Bubble	At least two types of reliability are reported that are appropriate[1] for the purpose of the tool, *and* the analyses are drawn from at least two samples representative of students across all performance levels, *and* the median of the estimates for each type met or exceeded 0.70.
Half Bubble	At least two types of reliability are reported that are appropriate[1] for the purpose of the tool, *and* (a) the analyses are drawn from at least one sample representative of students across all performance levels, *or* (b) the median of the estimates for each type met or exceeded 0.60.
Empty Bubble	Does not meet full or half bubble.

[1] Tests which require human judgment must report inter-rater reliability to be eligible for a Full or Half Bubble rating. Other types of reliability must include justification of appropriateness given the purpose of the tool.

2. Validity

Rating	Definition
Full Bubble	At least <u>two</u> types of <u>appropriately justified</u>[1] validity analyses are reported, ***and*** the analyses are drawn from at least one sample representative of students across all performance levels, ***and*** the median of the estimates for <u>both</u> types of validity met or exceeded <u>0.60</u> (or was within an acceptable range given the expected relationship with the criterion measure(s)).
Half Bubble	<u>One</u> type of <u>appropriately justified</u>[1] validity analysis is reported, ***and*** the analysis is drawn from a sample representative of students across all performance levels, ***and*** the median of the estimates met or exceeded <u>0.60</u> (or was within an acceptable range given the expected relationship with the criterion measure(s)).
Empty Bubble	Does not meet full or half bubble.

[1] Appropriately justified analyses must include criterion measures that are external to the progress monitoring system and theoretically linked to the underlying construct measured by the tool.

3. Bias Analysis Conducted

Bias analysis refers to an analysis that examines the degree to which a tool is or is not biased against subgroups (e.g., race/ethnicity, gender, socioeconomic status, students with disabilities, English language learners).

Rating	Definition
Yes	One or more of the following three types of analyses were conducted: 1. Multiple-group confirmatory factor models for categorical item responses 2. Explanatory group models such as multiple-indicators, multiple-causes (MIMIC) or explanatory Item Response Theory (IRT) with group predictors 3. Differential Item Functioning from Item Response Theory (DIF in IRT)
No	Fails "Yes."

Tools Chart Tab 2: Growth Standards

Note: For all standards in Tab 2, it is expected that evidence is drawn from a sample of students in need of behavioral intervention. Convincing evidence that children were in need of behavioral intervention may include one or more of the following: students have ED label; students are placed in an alternative school/classroom; students have demonstrated non-response to moderately intensive intervention (e.g., Tier 2); or students have demonstrated severe problem behaviors (e.g., Tier 3), according to an evidence-based tool (e.g., systematic screening tool or direct observation).

4. Sensitive to Behavior Change

Rating	Definition
Full Bubble	The basis for assuming that the data are sensitive to incremental change is strong (e.g., the range of possible scores is sufficient to detect small changes and documentation of sensitivity to change consistent with another criterion).
Half Bubble	The basis for assuming that the data are sensitive to incremental change is moderate (e.g., the range of possible scores is sufficient to detect a change and documentation of sensitivity to change).
Empty Bubble	Does not meet full or half bubble.
Dash	Data were not provided.

5. Reliability and Validity: Intensive Population

Rating	Definition
Full Bubble	At least <u>two</u> types of reliability are reported that meet the following criteria: - are <u>appropriate</u>[1] for the purpose of the tool, - are drawn from at least <u>two</u> samples representative of students in need of intensive intervention, and - the median of the estimates for each type met or exceeded <u>0.70</u>. *and* At least <u>two</u> types of validity are reported that meet the following criteria: - are <u>appropriately justified</u>[2], - are drawn from <u>at least one</u> sample representative of students in need of intensive intervention, and - the median of the estimates for both types of validity met or exceeded <u>0.60</u> (or was within an acceptable range given the expected relationship with the criterion measure(s)).
Half Bubble	One or more of the following were met: 1. At least two types of reliability are reported that meet the following criteria: are appropriate for the purpose of the tool, are drawn from at least two samples representative of students in need of intensive intervention, and the median of the estimates for each type met or exceeded 0.60. However, validity was not reported. 2. At least two types of validity are reported that meet the following criteria: are appropriately justified[1], are drawn from a sample representative of students in need of intensive intervention, and the median of the estimates for both types of validity met or exceeded 0.60 (or was within an acceptable range given the expected relationship with the criterion measure(s)). However, reliability was not reported. 3. One type of reliability was reported that met the criteria in #1 *and* one type of validity was reported that met the criteria in #2.
Empty Bubble	Does not meet full or half bubble.
Dash	Data were not provided.

[1] Tests which require human judgment <u>must</u> report inter-rater reliability to be eligible for a Full or Half Bubble rating. Other types of reliability must include justification of appropriateness given the purpose of the tool.

[2] Appropriately justified analyses must include criterion measures that are external to the progress monitoring system and theoretically linked to the underlying construct measured by the tool.

6. Decision Rules: Changing Intervention

Rating	Definition
Full Bubble	The data provided to support decisions about intervention change is <u>strong</u>, *and* based on analysis of progress monitoring measurement collected at least weekly over the period of time that is deemed necessary for the decision rules *and* from a sample of students that is in need of intensive intervention.
Half Bubble	The data provided to support decisions about intervention change is <u>moderate</u>, *and* is based on analysis of progress monitoring measurement collected at least weekly over the period of time that is deemed necessary for the decision rules *and* is from a sample of students that is in need of intensive intervention.
Empty Bubble	Does not meet full or half bubble.
Dash	Data were not provided.

7. Decision Rules: Choosing Intervention

Rating	Definition
Full Bubble	The data provided to support guidance on intervention choice (e.g., a class of relevant interventions or a specific intervention) are <u>strongly</u> evidence-based *and* based on analysis of progress monitoring measurement collected at least weekly over the period of time that is deemed necessary for the decision rules *and* from a sample of students that is in need of intensive intervention.
Half Bubble	The data provided to support guidance on intervention choice (e.g., a class of relevant interventions or a specific intervention) are <u>moderately</u> evidence-based *and* based on analysis of progress monitoring measurement collected at least weekly over the period of time that is deemed necessary for the decision rules *and* from a sample of students that is in need of intensive intervention.
Empty Bubble	Does not meet full or half bubble.
Dash	Data were not provided.