
Component IV: Comprehensive and Balanced Assessment

Good decision making for groups and individuals requires good information. This critical component acknowledges the central role of assessment in effective MTSS-RTII.

Operating Assumptions:	Specifically refer to Guiding Principles #4, 7, and 8.
Key Definitions:	Refer to the Glossary for the following terms: <i>Balanced & Comprehensive Assessment System</i> <i>Benchmark/Periodic Progress Monitoring</i> <i>Progress Monitoring</i> <i>Formative</i> <i>Ongoing Progress Monitoring</i> <i>Screening</i>

The Significance of a Comprehensive and Balanced Assessment System

We have known for decades that students' academic success starts with committed, knowledgeable educators and the highest quality classroom instruction, informed by research, and supported by a standards-based curriculum (IES Practice Guides for Reading and Mathematics, 2009a, 2009b; Readance & Barone, 1997; Riordon & Noyce, 2001; Rosenshine, 2012; Steedly, Dragoo, Arefeh & Luke, 2008; Williams et al., 2005). In addition, we know that improving students' social and behavior functioning often improves academic performance and vice versa. As with academic success, this requires a system of explicit and responsive teaching of behavioral expectations leading to potential gains in positive conduct, academic performance, and teacher-student interactions and reductions in office referrals and suspensions (Bui, Quirk, Almazan & Valenti, 2010; Rimm-Kaufman et al., 2012). Recommendations for improving outcomes for students who are struggling in the area of behavior are, therefore, intertwined with improved instructional contexts (IES, 2008, 2009).

While contextual factors matter, there has always been considerable variability in student outcomes across schools—even when the contexts are quite similar. Historically, educational research and practice have focused attention first on how to change struggling students. Current emphasis, however, is on first ensuring that students experience the highest-quality instruction – instruction that is differentiated and responsive to diverse students and provides appropriate support for both accelerated academic success and socially effective behavior.

COMPREHENSIVE AND BALANCED ASSESSMENT

Purposes and Tools

A well-designed balanced assessment system includes tools and processes that are effective to address various assessment purposes. These tools are often called: *screening, diagnostics, formative progress monitoring, interim/benchmark progress monitoring*, and summative or *outcome assessment*. These types of assessment tools are employed to address five purposes of assessment: *planning learning, supporting learning, monitoring learning, verifying learning*, and *investigating the cause of learning difficulties*. See Table 5 for an example of assessment types organized by purpose.

School districts and/or supervisory unions can use the concept of a *balanced assessment system* to identify and organize the assessments they use by *purpose*. An analysis of a particular school's overall assessment system may help a school to identify whether or not some types of assessments are used more frequently or receive more emphasis over other types. If the system appears out of balance, adjustments should be made. The specific tools and processes may differ across districts and schools, but a trustworthy system is comprehensive enough to address all purposes and to capture the full range of critical components within the academic or behavioral domain.

Utility

Assessment information alone is useless; it should invite action. Educators must be wise consumers of assessment data, understanding its properties and appropriate uses. The data and information provided by assessment must be examined, discussed, reflected upon, and used to make decisions. As the consequences of these decisions become more serious, so must the range and quality of the assessment information. Overall quality, accuracy and timeliness affect the utility of assessment data, but so do systems' factors such as the school's capacity to analyze and interpret data and the school's structures for supporting collaborative discussion and widespread use of the information (Lachat & Smith, 2005).

To use assessment data wisely, educators should understand the multiple components of reading and math across developmental levels and as they are implicated in diverse content. Educators at different grade levels may gather different types and amounts of assessment data but they are, nevertheless, each responsible for understanding and using information to improve instruction and learning for all students. It is also important to note in Table 5 that different types of assessment can be used for more than one purpose and, generally, no one piece of assessment information can fulfill all purposes.

COMPREHENSIVE AND BALANCED ASSESSMENT

Table 5. Balanced Assessment System By Purpose

PURPOSE	WHAT DOES THIS LOOK LIKE?	ASSESSMENT OPTIONS	NOTATIONS
<p>Screening</p> <p>To Identify Students Who Require a Closer Look</p>	<p>Data that:</p> <ul style="list-style-type: none"> Identify or flag students who are struggling or may be at-risk of school failure and who will require closer monitoring Raise unanswered questions about individuals or groups of students Or, the effectiveness of core academic and behavioral curricula 	<ul style="list-style-type: none"> Dedicated screening tool Formal review of existing progress monitoring data On-going formative assessment data 	<ul style="list-style-type: none"> Data for screening purposes are collected for all students one or more times a year. Tests dedicated to screening (sometimes called universal screeners) are generally most important when: 1) there is no comprehensive assessment system in place that provides on-going information about individual students or, 2) students are new to school (i.e. PreK-K and/or middle/high school) and/or there are many new students each year. Screening for behavioral concerns involves reviewing trends for individual students as well as relevant themes within the total or disaggregated population (i.e., grade level/class, problematic time of day, location, etc.). If behavioral data are reviewed on a routinely fixed schedule, they may serve the purpose of screening in the absence of a standardized tool.

COMPREHENSIVE AND BALANCED ASSESSMENT

Table 5. Balanced Assessment System By Purpose (cont'd)

PURPOSE	WHAT DOES THIS LOOK LIKE?	ASSESSMENT OPTIONS	NOTATIONS
<p>Diagnostic</p> <p>To Investigate and Analyze Learning Difficulties</p>	<p>Data that:</p> <ul style="list-style-type: none"> • Inform the educator about possible causes of student difficulties • Explore the domain (literacy, mathematics or behavior) more comprehensively • Identify appropriate focus for instruction/intervention • Explore and identify possible effective instructional/intervention approaches 	<ul style="list-style-type: none"> • Standardized diagnostic assessment tools • Closer and more detailed analysis of existing progress monitoring data • Additional measures/data to get a more comprehensive picture • Observations, interviews, and work samples 	<ul style="list-style-type: none"> • Diagnostic assessment is conducted with only some students but is often necessary to plan instruction and/or intervention to meet the needs of students who are experiencing difficulty. • The goal is to plan more effective and tailored instruction and/or intervention based on more refined information. • Students often provide excellent insights into their own learning strengths and needs. Their self-assessments should be carefully considered.
<p>Progress Monitoring: Formative</p> <p>To Inform Instruction</p>	<p>Data that:</p> <ul style="list-style-type: none"> • Provide information to both educators and students about what has been learned, which objectives have been addressed, and what techniques have been successful • Help educators make decisions about what to teach, how to adjust their instruction along the way, and/or where to start • Data that reveals depth of understanding and partial or developing understandings 	<ul style="list-style-type: none"> • Any data that shows teachers what has been learned and what needs to be addressed instructionally • Student engagement in the process is pivotal 	<ul style="list-style-type: none"> • Standardized information can be very helpful in planning overall instruction for groups of students. • As well, educators use on-going formative assessment data (including student self-assessment) to refine and adapt instruction for groups and individuals (see note above about student self-assessment).

COMPREHENSIVE AND BALANCED ASSESSMENT

Table 5. Balanced Assessment System By Purpose (cont'd)

PURPOSE	WHAT DOES THIS LOOK LIKE?	ASSESSMENT OPTIONS	NOTATIONS
<p>Progress Monitoring: Periodic Benchmarking</p> <p>To Monitor Progress</p>	<p>Data that:</p> <ul style="list-style-type: none"> • Show educators (and others) what progress has been made during a specific period • Track student progress on identified tasks/benchmarks 	<ul style="list-style-type: none"> • On-going formative progress monitoring data • Interim/periodic benchmark assessments • Standardized outcome measures 	<p>An array of data can and should be used to monitor student progress.</p> <p>A robust progress monitoring system can function in place of a separate screening measure.</p>
<p>Outcome or Summative</p> <p>To Verify Learning</p>	<p>Data that:</p> <ul style="list-style-type: none"> • Confirm what students know and can do; typically at the end of year, semester, course, or instructional unit • Reflect an appropriate and comprehensive picture of the domain (literacy, mathematics, behavior) 	<ul style="list-style-type: none"> • Standardized test data to assess outcomes • Benchmark progress monitoring data • Formative assessment data demonstrating learning 	<p>Because data provide information about individual students and also about groups, it can be used to make decisions about instruction, curriculum and program adjustments.</p> <p>Protocols for examining outcome data should support educators as they use data for diverse purposes.</p>

COMPREHENSIVE AND BALANCED ASSESSMENT

Assessment Quality

The quality of assessments matters. Since MTSS for RtII rests so significantly on the decisions made from assessment information, that information must be both **trustworthy** and **useful**. This is more complicated than it may seem since the trustworthiness and utility of assessment information depend on many things – including the assessments themselves (the tests, activities, and content), variation in administration and/or scoring of assessments, as well as the uses to which the assessment data are put.

Two key concepts for judging the quality of an assessment relate to *reliability* and *validity*. Although these criteria were developed for use in judging norm-referenced tests, the underlying concepts have implications for other assessment measures (such as standards-based assessments) as well.

Reliability refers to the degree to which an assessment measure produces consistent and accurate results over time (e.g., a test of reading comprehension has been pre-determined by test developers to produce consistent results for students). Efforts should always be made to make sure that any variation in the data is the result of changes in student performance – not changes in administration, interpretation, random factors, etc.

Validity is judged on the meaningfulness and usefulness of the interpretations made by users of the assessment results (Standards for Educational and Psychological Testing, 1999). For the past 15 years, validity has been considered, not as an attribute of the assessment itself, but rather in relation to the testing purpose and to the quality of inferences drawn from the assessment. Presently, there is considerable debate about the appropriate way(s) to determine and judge validity (Gorin, 2007; Lissitz & Samuelsen, 2007), with some scholars arguing that “construct validity” (judgments about the construction of the test) should hold more sway. In either case, a major factor in the valid use of tests is the extent to which there is a match between the test and the test user’s concept of the domain being measured. In our case, we would ask whether the test content and process matches our concepts of reading and/or mathematics and/or behavior so that we can make good inferences about students within a specific setting.

In the context of RtII, validity issues are critical. When we interpret the results of any assessment(s), we should ask not only what the data suggest about groups of students but also what they say about the systems surrounding the students’ learning/behavior. When the data indicate that a great many students require intervention, educators must consider whether the curriculum and instruction need attention.

COMPREHENSIVE AND BALANCED ASSESSMENT

A particular validity concern surrounds screening measures. Assessment data are only useful for screening if they can demonstrate *predictive validity*. That is, the results from a screening measure must predict students' performance on (other) measures of interest (math, reading, behavior). If there are too many “false positives” (students are identified as having a problem when they do not) or “false negatives” (students are not identified when they should be), the validity of the assessment is called into question. This problem is quite complex and results for most available “screening measures” are vulnerable on this count, leading Hosp, Hosp, and Dole (2011) to conclude that, “the use of a single measure is not prudent for screening decisions” (p. 129).

Finally, not all tests are appropriate for all interpretations and uses. We need to always ask ourselves whether the conclusions drawn from an assessment are valid in relation to the thing(s) we are interested in measuring. In addition, educators must always be asking themselves if their data are trustworthy – even if the assessment tool or approach is not subject to formal tests of reliability.

The quality of assessment measures becomes more important as the stakes increase. It may be appropriate to use “teacher-validated” formative assessments to plan and adapt daily and weekly instruction. However, the process used to determine whether or not a particular intervention is working or whether or not a student is eligible to receive special education services requires additional consideration regarding the degree to which an intervention and/or assessment measure is reliable and trustworthy. Generally speaking, this will require multiple measures – not just repeated use of one measure, but different sources of assessment data.

Standardized and commercially produced assessment measures should contain information to help in the selection of tests. Teachers and school teams, however, will also need to determine what measures are valid in the context of that school. Factors that might be considered include the school's curriculum, student population, cultural considerations, and teacher perceptions of what constitutes valid assessment practices.

Finally, it is important to remember that “assessment” does not necessarily mean “test.” A great deal of time, attention, and money has been spent in recent years to improve the quality of high-stakes tests. But Stiggins (2007) reminds us that, behind these considerable accomplishments, there is almost complete neglect of assessment where it exerts the greatest influence on pupils' academic lives: day to day in the classroom, where it can be used to help them learn more” (p. 10). Stiggins also suggests:

The principal assessment challenge that we face in schools today is to ensure that sound assessment practices permeate every classroom—that assessments are used to benefit pupils. This challenge has remained unmet for decades, and the time has come to conquer this final assessment frontier: the effective use of formative assessment to support learning (p. 10).

These concerns are especially critical at middle and high school levels where the range and utility of conventional tests are so much smaller (Torgesen & Miller, 2009).

COMPREHENSIVE AND BALANCED ASSESSMENT

Essential Elements of Assessment in MTSS-RtII

Layered Approach

A multi-tiered system of RtII requires a comprehensive and balanced assessment system. A balanced assessment system is one that may be “bigger” than the RtII framework in that it articulates and involves all aspects of assessment that occur in a particular school district/supervisory union over the course of a given year and over years. Efficient RtII assessment systems often involve a layered approach in which screening techniques are first used, both to identify which students require further (diagnostic) assessment and to provide aggregate data about the nature of student achievement overall. More detailed diagnostic assessments, used with some students, can provide more comprehensive data to inform instruction.

Although most approaches to RtII focus on screening and progress monitoring, effective multi-tiered systems also contain processes, procedures, and measures for gathering and analyzing diagnostic data. Ideally, the initial review by a school/district of its existing assessment system will reveal measures and tools used for each purpose. However, some schools may find that they need additional measures for a balanced and comprehensive approach.

Although it is much less common, assessment of the instructional context has long been acknowledged as essential in determining a good match between learning and instruction (Lipson & Wixson, 1986, 2012). Recent attention to behavioral outcomes has focused more attention on contextual issues. Because the quality of the instructional context is important for both academic and behavioral achievement, we provide a tool to assist districts/schools in considering how they assess critical contextual features as well as student outcomes (see Table 6).

Table 6. Comprehensive and Balanced Assessment System ⁵

Grades:	Component Areas of Math, Literacy or Behavior					Contextual Features		
Core Curriculum All Students			Outcomes/Screening District/Classroom Progress Monitoring				Classroom Practice(s)	
Targeted, Strategic Interventions Some Students			Diagnostic Intervention/Progress Monitoring				Features of Intervention and Responsive Instruction	
Intensive Interventions Few Students			Diagnostic Intervention/Progress Monitoring				Specific and Tailored Modifications	

⁵ Adapted from Rhode Island RTI Initiative (2007). Building A Comprehensive Assessment System: The Role of Assessment in RTI (Module #3). Rhode Island Technical Assistance Project (RITAP). Accessed at http://www.ritap.org/rti/content/modules/1_BuildingAComprehensive%20AssessmenSystem.POWERPOINT.ppt.

COMPREHENSIVE AND BALANCED ASSESSMENT

Screening

Data used for *screening purposes* are typically collected two to three times per year. All students in a school and/or at a particular grade level participate in one or more common assessments of academic and/or behavioral achievement. Because data are collected on all students, it is important that the system work efficiently. Some schools/districts identify distinct “screening” instruments that are short and easy to administer. However, it is possible within a comprehensive assessment system to use periodic benchmark or interim progress monitoring data for the purpose of screening. When there are common assessments created or agreed upon within a school, even formative assessments can serve the screening function.

The screening function is most critical at points of transition (i.e., entry into kindergarten, movement to middle or high school) when the body of information about an individual student may be small or may be less accessible. Most schools have some type of kindergarten screen, but it should be examined to ensure that key information is collected. District-wide comprehensive assessment systems often neglect, however, procedures for aggregating existing information of students for use in screening students as they transition to middle or high school. Yet, there is typically a great deal of information that can be used. Often educators consider a wide range of available data (e.g., standards-based report cards, curriculum based measures, district level standardized measures, etc.) as they make screening decisions that may help teachers and/or teams to establish a baseline from which to set goals for targeted interventions for students receiving support at the Tier II and III levels.

Tools used in universal screening are generally not designed to collect enough diagnostic information to plan and monitor instruction. However, they should be tools that are reliable and valid in terms of their ability to identify students who may need to be considered for additional support within the RtII framework or who require a closer look using other assessment(s).

Data collected for screening purposes needs to be analyzed to determine whether each student is (or is likely) to meet, exceed, or not meet benchmarks. Ideally, a screening measure will identify only those students who are really at risk (true positives) and would never flag a capable student as struggling (false positive). Educators must be wary of using a single screening tool since virtually no available assessment meets this standard and many are quite likely to misidentify students one way or the other.

Progress Monitoring

Within the RtII framework, both types of progress monitoring (formative and interim/benchmark) are critical to decisions about the provision of instruction and intervention at all levels of a multi-tiered system. Both types help teachers and teams to provide responsive instruction and intervention and are used to decide whether or not planned interventions are working. Tools used in progress monitoring may vary. Here too, schools need to review their assessment systems to identify which assessment procedures are best suited to progress monitoring.

Whole class progress monitoring may be conducted through formative assessment using teacher-designed assessments and/or standardized quantifiable measures. Progress monitoring for students receiving Tier II and III interventions may also be conducted with a variety of measures. Depending on the assessment schedule in the school/district, and the types of assessment data available, some students (at Tiers II and III) may be monitored more closely using additional or different assessments, as they will be used to make more high stakes decisions.

COMPREHENSIVE AND BALANCED ASSESSMENT

Progress monitoring data are also critical to the special education eligibility process for those students who may be considered for eligibility under the category of specific learning disabilities. Schools choosing to adopt the MTSS-RtII framework as a replacement for using severe discrepancy models and other formulas for determining the presence of a learning disability should have a carefully considered plan. That said, certain types of progress monitoring data might be used by any school as part of the determination of “adverse effect” in a comprehensive evaluation, regardless of the type of disability under consideration.

Diagnostic Measures

Assessment for *diagnostic* purposes is a critical part of a school/district’s MTSS-RtII framework. As indicated above, few universal screening tools and only some progress monitoring tools are designed to probe student learning at a deeper diagnostic level. As such, those engaged in planning and implementing an approach to assessment within the RtII framework will need to consider what tools are appropriate and necessary for a more diagnostic approach to designing, implementing and evaluating instruction and intervention.

Practical Matters

A successful comprehensive assessment system should invite action. The system must include not only what assessments will be used but also address:

- *Where* will the data be stored/collected?
- *Who* will review the results of assessment(s)?
- *When* will these data be considered?
- *How* they will be used?

Examining data

For example, in a school conducting assessments of all students three times yearly, all teachers would review the data for their own class and grade and teams would examine data across a grade or for the school. However, one or more teams (e.g., the Educational Support Team (EST), grade level team, or a team associated with a Professional Learning Community (PLC)) may review data for some students more often. Classroom teachers use the data to adjust Tier I curriculum and instruction, and/or to identify students who may be in need of further assessment or more intensive levels of support or intervention. Specialist teachers use the data for these same purposes, but also to determine whether the intervention(s) being used are effective for individual students, whether more assessment data are needed, and/or whether decisions about eligibility are suggested.

Monitoring progress for some students

A similar set of questions and a system for making decisions is essential for students receiving interventions. Teachers and/or support personnel who are responsible for implementing interventions are generally assigned to collect progress monitoring data, and often teams (i.e., PCLs, ESTs or other teams designated as “RtI” teams) will review the results of progress monitoring on a pre-determined basis to identify whether or not interventions need to be changed, intensified, decreased, or discontinued.

COMPREHENSIVE AND BALANCED ASSESSMENT

On-going formative assessment as well as more frequent periodic benchmark assessment should be used with vulnerable students. Teachers and specialists who are providing interventions should examine student performance and their own instruction on a daily/weekly basis and use these data to make adjustments. This is especially true for very young students (K-2). Although there are no “hard and fast” rules regarding the timing and review of progress monitoring, schools engaging with the MTSS-RtII framework should consider and agree upon a set of guidelines or “decision rules” for progress monitoring as part of their overall design process.

Making decisions about interventions

How long should a single intervention be continued before it is terminated and/or before some other approach is taken? The literature is mixed on this question and is available only for elementary grade levels. The consensus agreement among professionals using the existing literature is that progress monitoring at eight to ten weeks should reveal progress and, if it does not, school teams and teachers need to review the data carefully to identify potential change(s). The nature of the data to be used to make decisions is also critical here. As we noted above, professionals should be examining on-going (formative progress monitoring) data in a continuous way. The frequency for administering more formal (benchmark progress monitoring) assessments is open to discussion. Although some researchers have suggested that weekly testing is required, the most recent research indicates that brief weekly tests (e.g. one-minute tests of oral reading and/or computation) may not be necessary since little additional information is added before 8-10 weeks (Jenkins, Graff, & Maglioretti, 2009).

Because data-based decision-making is a key element of any assessment system, schools choosing to adopt an MTSS-RtII framework will need to consider the capacity of teachers and decision-making teams to organize, review, analyze and make individual and school wide data-based decisions. Any system of assessment is only as good as the ability of its consumers to understand and use the data that are generated through that system, and for many teachers and teams, data-based decision-making is a skill that needs to be taught and supported on an ongoing basis.