“Understanding Types of Assessment Within an RTI Framework”

**Slide 1:** Welcome to the webinar “Understanding Types of Assessment Within an RTI Framework.” This is one of 11 webinars that we have developed at the National Center on Response to Intervention (NCRTI). This webinar focuses on the types of assessments that are commonly used within an RTI framework. If you have not already, you may want to print out the accompanying handouts. We recommend that you pause the webinar periodically to complete those handouts. My name is Dr. Rebecca Zumeta, and I coordinate technical assistance and product development for the National Center on Response to Intervention.

**Slide 2:** The Center has developed a series of webinars to cover information that is important for early implementers. On this slide you can see the variety of webinars that are available in the series. While you can watch the webinars in any sequence, we recommend that you first start by watching the “What is RTI?” webinar in order to ensure you have an understanding of how RTI is discussed in this series and the language that we use. Following this webinar, you can learn more about assessments by viewing the webinars focused on screening and progress monitoring.

**Slide 3:** Upon completion of today’s webinar, participants will not only be able to identify when to use formative, summative, and diagnostic assessments, but will also understand the differences between norm-referenced and criterion-referenced assessments, and also recognize the benefits and drawbacks of general outcome measures and mastery measures for progress monitoring.

**Slide 4:** One helpful way to check your understanding throughout this webinar is to complete the vocabulary handout that we have included in the handouts for this session. This handout provides a table with a list of key terms on the left side, and with columns on the right side for you to track your understanding before and after viewing the webinar. Before viewing the webinar, you should make predictions of the meanings of each term. Then, throughout the webinar, you will complete the final meaning column, based on the definitions provided, along with a picture/sketch/example column, where you can add anything helpful for remembering what you have learned. You can see how in this example, I was able to clarify the meaning of primary prevention level after watching the webinar.

If you have not already done so it may be a good idea for you to pause the webinar to make predictions; press play when you are ready to continue.
Slide 5: There are three types of assessments commonly used within an RTI framework. These are summative, diagnostic, and formative assessments. Summative assessments occur after instruction has occurred, and are assessments of what has been learned. Diagnostic assessments occur before instruction takes place to help identify different skill strengths and weaknesses. Formative assessments occur during instruction, and are intended to help inform instruction.

We will now spend some more time on each type of assessment.

Slide 6: The first type of assessment is summative. These measure what students learned over a period of time. They are typically administered after instruction, and can help to determine what to teach, but not necessarily how to teach. Summative assessments are typically administered to all students and are often used for accountability decisions and to determine whether students have mastered certain skills. While summative assessments can also be used to allocate resources, because they occur after instruction, this decision would be considered reactive because it is occurring in response to how students performed.

If the only assessments you use are summative, you will need to wait until the end of the school year to make instructional decisions, and that may not always be efficient. So if your goal is to prevent poor learning outcomes, not just identify them at the end of the year, you may want to consider using additional assessments in addition to summative assessments.

Slide 7: One category of summative assessments is high-stakes tests. This includes state assessments or state achievement tests, the Graduate Record Exam (GRE), ACT, SAT, Praxis Tests, and other final course exams.

Although they provide useful information, they usually require time away from instruction are not always valid for individual decision making or identifying individual students strengths and weaknesses, and they don’t tell us how or why students achieved different scores.

Slide 8: Diagnostic assessments, on the other hand, are measures of students’ current knowledge and skills, and are typically used before instruction to identify strengths and weaknesses and identify a suitable program of learning. They are again administered before instruction occurs, and help teachers to determine whether or not a particular instructional program or intervention program may be effective. Because diagnostic assessments provide detailed information about individual student learning, they are often most effective in understanding the needs of individual students, rather than all students. So they are therefore typically administered to only some students, or to students who’ve been identified as being at risk for learning problems. Diagnostic assessments can help teachers determine what to teach and help in the selection of appropriate interventions. They provide useful information, but should be used sparingly, as they can be very expensive and time consuming.
Slide 9: Some examples of diagnostic assessments include Qualitative Reading Inventory, Diagnostic Reading Assessment, Key Math, and Running Records with Error Analysis, among many others.

Slide 10: The third type of assessment that is commonly seen in an RTI framework is formative assessment. And this is a form of evaluation used to plan instruction in a recursive way. These assessments tell us how well students are responding to what we are teaching. With formative assessment, student progress is systematically assessed during the school year, during instruction, to help provide teachers with continuous feedback about learning successes and failures. Formative assessments are typically administered to all students during benchmarking (screening), and then to some students within the context of progress monitoring.

Formative assessments may be informal or formal. Informal assessments are not data driven but rather content and performance driven. Examples may include things such as observations and teacher-made assessments. Formal assessments, which have data that support the conclusions made from the test, including evidence of reliability and validity, are more common within the RTI framework. These are typically referred to as standardized measures. These tests have been tried before on students and have statistics that support the conclusion that when a student is responding or performing well on the assessment that they are not at risk and that students performing below certain cut scores or point on the assessment are at risk. The data are often mathematically computed and summarized. Scores such as percentiles, stanines, or standard scores are commonly received as part of this type of assessment.

Some examples of more formal formative assessments are curriculum-based measures, curriculum-based assessments, pre/post tests, and benchmark assessments. Some less formal types may be teacher-made quizzes, or exit forms that teachers may require of students as they complete a lesson at the end of each day.

Slide 11: Formative assessments allow teachers to diagnose skill, ability, and knowledge gaps, and measure progress; and evaluate instruction. There are a variety of educational decisions that can be made from the use of formative assessments, including the identification of individual students who are nonresponsive to instruction or interventions. They can also be used to drive curriculum and instructional decisions, to evaluate whether a particular instructional program is working, and to compare the quality of different instructional or intervention programs. They may also be used to help drive decisions about resource allocation.

Formative assessments are not necessarily used for grading purposes. However, they are again intended to help teachers determine whether or not their instruction is effective and whether changes are needed.
Slide 12: There are two types of formal formative assessments that are typically used. One is a mastery measure that is used to index a student’s successive mastery of a hierarchy of objectives and skills; these tend to be curriculum dependent and are reliant on the order in which skills have been taught in a particular curriculum. General outcome measures (GOMs), on the other hand, reflect the overall competence of a student in the annual curriculum and the growth of student competence in skills within the annual curriculum. Examples of general outcome measures include AIMSweb’s Reading-CBM, Early Literacy and Numeracy, Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Retell or Oral Reading Fluency, and iSTEEP’s Oral Reading Fluency, among others.

Slide 13: The following quote is often helpful in thinking about the differences between summative and formative:

“When the cook tastes the soup, that's formative. When the guests taste the soup, that's summative.”

Slide 14: At this time we are going to complete the Types of Assessment handout. For each of the first four questions, there are three possible answers (summative, formative, or diagnostic). Take a moment and pause the webinar to match the type of assessment with the description that is listed. Once you have completed the handout, press play to resume the webinar.

Slide 15: The correct answers to this handout are as follows:

In the first question, the description measures a student’s current knowledge and skills for the purpose of identifying a suitable program of learning. That is a diagnostic assessment.

The second item tells us what students learned over a period of time (in the past). It may tell us what to teach but not how to teach. That is a summative assessment.

The third item tells us how well students are responding to instruction. That would be a formative assessment.

The next set of questions asks when assessments should be administered. Assessments administered during instruction are formative assessments. Assessments administered before instruction are diagnostic assessments. Assessments administered after instruction are summative assessments.

The next set of questions asks who the assessments are typically administered to. Assessments of all students during benchmarking/universal screening and some students for progress monitoring would be considered formative assessment. Assessments administered to all students are summative assessment. Assessments administered to some students are diagnostic assessments.
For the purpose of driving educational decisions, assessments that are used for accountability for meeting standards or desired outcomes, for assessing skill mastery by students, and for future allocation of resources based on outcomes, we would typically use a summative assessment. For identification of students who are nonresponsive to instruction or interventions, determination of curriculum and instructional decisions or program evaluation, resource allocation/alignment to meet student needs, or comparison of instruction and intervention efficacy, we may use formative assessment. For help determining what to teach and help in selecting an intervention, we may use a diagnostic assessment.

**Slide 16:** Now we will spend some time talking about the difference between norm-referenced and criterion-referenced assessments.

Norm-referenced test include tests like the ACT, SAT, and GRE. Students’ scores are compared with each other using percentile scores. For example, a student who scored in the 64th percentile scored better than 64 out of 100 students who took the test. He or she scored lower than 36 out of 100 students who took the test.

Criterion-referenced tests are often standards-based tests that compare a student’s performance to a criterion for mastery of a skill or objective. The score indicates whether the student met the mastery criteria, and so it often looks like a pass/fail or proficiency score.

Assessments, including screening assessments, can be both norm referenced and criterion referenced depending on what information you are trying to gather. If you are interested in knowing how a student compares with other students in the screening process, you would probably be most interested in a norm-referenced assessment. If you want to screen for which students met or did not meet a set criterion, typically a score, then you would be interested in a criterion-referenced assessment. This will be discussed more later.

**Slide 17:** We are now going to discuss in more detail the two common formats for formative assessments. Each type of measure has value, although they are typically used for different reasons.

Mastery measurement indexes a student’s successive mastery of a hierarchy of objectives or skills, whereas general outcome measures reflect overall competence or skill in an annual curriculum.

**Slide 18:** A mastery measure describes mastery of a series of short-term instructional objectives. To implement mastery measurement, a teacher will typically determine a sensible instructional sequence for the school year and design criterion-referenced testing procedures to match each step in that instructional sequence.
Slide 19: Here is an example of a typical fourth-grade math computation curriculum, with 10 objectives the teacher plans to accomplish for the year. The first objective is multi-digit addition with regrouping.

Slide 20: While teaching multi-digit addition with regrouping, the teacher may give assessments that look something like this. There are 10 problems, all dealing with multi-digit addition with regrouping.

Slide 21: As the teacher gives this assessment, he or she could graph the student’s results. In the first week of teaching the objective, the student did not get many answers correct, as you can see on this graph, but as you can also see, as the weeks go by, the student gets better, until he or she has had three consecutive weeks of getting 80 percent or more of the problems correct. Once this happens, then the teacher is confident that the student has mastered the skill and moves on to the next objective.

Slide 22: The next objective in this curriculum is multi-digit subtraction with regrouping.

Slide 23: Again, the teacher is teaching this objective and while doing so he or she may give assessments that look like this: 10 problems all dealing with multi-digit subtraction with regrouping. Note that there is no practice of multi-digit addition with regrouping, which was the previous skill that was taught.

Slide 24: Again, the student’s progress is charted. A problem with this chart is that you cannot tell if the student is learning the objectives at a pace fast enough to allow him or her to learn all of the objectives in the curriculum by the end of the school year.

Slide 25: However, there are some important advantages to mastery measures as well. For example, by focusing on a single skill, teachers can assess whether a student can learn certain target skills in isolation. They can also use the information from the ongoing data to make decisions about how to change instruction for target skills. Until recently, the psychometric properties of most mastery measures were not established, and many were not valid and reliable. However, if you spend some time looking at the Mastery Measures section of the Progress Monitoring Tool Chart, there is increasing research demonstrating the validity and reliability of some of these tools. Mastery measures are typically not valid as screening measures; however, they are often used to monitor progress for students identified through screening as being potentially at risk.

Slide 26: At the same time there are some challenges associated with mastery measurement. First the hierarchy of skills taught is often logical, not empirical. This means that while it may seem logical to teach addition before subtraction, there is no evidence-based research that says you have to do it that way. Second, the assessment does not reflect maintenance or generalization
of skills. For example, after teaching subtraction with regrouping, you don’t know if the student remembers how to do addition with regrouping. The number of objectives mastered does not necessarily relate well to performance on criterion measures. This means how a student does on these isolated skill assessments does not indicate how he or she will do on an end-of-year standardized assessment in math. Measurement methods are often designed by teachers, and therefore they have unknown reliability and validity. So again, if a teacher develops his or her own assessment we don’t know how that assessment may compare to an assessment that another teacher may develop, or if this is a reliable or consistent measure of student skills. Also, we cannot compare scores longitudinally; you will not be able to compare a student’s score on a mastery measure in September with his or her score on that measure in May, because different skills are being assessed—so the scores don’t mean the same thing.

**Slide 27:** General outcome measures, on the other hand, address many of the problems associated with mastery measures. First, they are program independent, meaning they reflect overall competence in the yearlong curriculum as opposed to being dependent on a particular program. They also describe an individual child’s growth and development over time. Third, they provide a decision-making model for designing and evaluating interventions. And finally, they can be used for individual children and for groups of children, and for this reason can be used for both screening and progress monitoring.

**Slide 28:** Some common characteristics of general outcome measures are that they are simple and efficient, which means that they are brief assessments that typically take no more than 1–10 minutes to administer and they provide you with enough information to be able to make quick decisions. Classification accuracy can be established. That means that the classification accuracy can tell you the extent to which a screening tool is able to accurately classify students into “at risk” and “not at risk” categories. They are sensitive to improvement. That means that they are able to measure the extent to which improvement on this assessment actually indicates that a student is making progress toward the outcome of interest, and whether the student is making a level of progress that is suggestive of him or her being at risk of a problem in that academic content area. They also provide performance data to guide and inform a variety of educational decisions, and importantly, the norms that are used at either the national or local level can be used for comparisons across time.

**Slide 29:** The advantages of GOMs include the focus on repeated measures of performance, so that there are no assumptions about instructional hierarchy for determining measurement. In other words, they can be used with any instructional approach. Also, general outcome measures incorporate automatic tests of retention and generalization, because they assess the students on skills they have already been taught, and may require a student to recall what they’ve learned; and they also present students with skills that they have not yet been taught, to see if they have
the ability to generalize or extrapolate what they learned to the next step of the instructional sequence. This allows the teacher to constantly assess whether the student is retaining what was taught earlier in the year and whether the student is able to think about what steps may come next.

The point here is not to say that practitioners should always use general outcome measures or mastery measures, but that it is important to think about which measure is better for each objective.

For screening purposes, general outcome measures should always be used, because GOMs work better to fulfill the objective of a screener instrument.

**Slide 30:** For an example of a general outcome measure, you can think about curriculum-based measurement (CBM).

A curriculum-based measure is a general outcome measure of a student’s performance in either basic academic skills or content knowledge. CBM was initially developed more than 30 years ago by Stan Deno and others at the University of Minnesota Institute for Research on Learning Disabilities, and is a reliable and valid measurement system for evaluating basic skills growth.

Research began in the area of reading and has expanded to additional tools in the areas of literacy, mathematics, and writing, as well as tools in Spanish.

**Slide 31:** One example of a common CBM is an oral reading fluency assessment. Students would be asked to read this passage aloud while being timed for one minute. The score would be based on the number of words read correctly per minute.

**Slide 32:** In summary, both types of formative assessment provide useful, but different, information. Mastery measurement data indicate progress toward mastery of specific set of subskills (for example, decoding short vowel sounds, or calculating nine math facts), which is indicated by the graph on the left of the slide. On the other hand, GOMs indicate progress toward a broader outcome. So that approach involves the application of all of the subskills learned over time; this is demonstrated by the graph on the right side of the slide, which shows a student’s progress over time compared to an instructional goal that was set for him or her.

One key difference between mastery measures and general outcome measures is comparability of data of time. With a GOM you can compare the score a student received in May to a score he or she had in September. This cannot be done with mastery measures, because the skills assessed by mastery measures differ over time, whereas the skills assessed by GOMs remain constant over time.
Slide 33: Thank you for taking the time to listen to “Understanding Types of Assessments Within an RTI Framework.” To find more resources on this topic or to view other webinars in the implementer series, visit www.rti4success.org. Here you can also view additional information from the RTI Action Network and the IDEA Partnership.