"Making Sound RTI Decisions: Who Responds to Secondary Prevention"– Webinar

Slide 1: Welcome to the National Center on Response to Intervention webinar. Today we're going to be talking about making sound RTI decisions and we're going to be talking about who responds to secondary prevention, which is sometimes referred to as "Tier 2."

Slide 2: So like I said, in this session we're focusing on RTI, which is a multi-level prevention system. Today I'll be referring to the first two levels of the prevention system. The first level is primary prevention, where the classroom teacher delivers the classroom instructional program while implementing adaptations and accommodations to address individual student needs. Average students also receive secondary prevention and in most RTI models secondary prevention occurs in the form of small group tutoring. One key decision point within RTI is to determine which students do and do not respond to secondary prevention. In making these decisions we want to be as sure as possible that the children we designate as responders will in fact thrive in a general education program after tutoring ends. And we also want to be as sure as possible that the students who are designated as non-responders in fact do really require ongoing and more intensive intervention to support their adequate learning. In this webinar I'm going to discuss what research can tells us about methods for distinguishing responders from non-responders.

Slide 3: So why is it important to quantify a response using valid decision roles? Well the first reason is that in many RTI systems some students linger in secondary prevention. This lingering in secondary prevention is not legal as discussed in a US Department of Education memo on January 21st, 2011. In that memo, Melody Musgrove stated that: "An RTI process cannot be used to delay or deny an evaluation for eligibility under IDEA." She spoke about how students with disabilities must be identified in a timely manner. The thrust of the memo was that long term secondary prevention cannot be used as an alternative for providing services without an IEP to a student with a disability. It's also ineffective for students to linger in secondary prevention. If a student has not responded to secondary prevention within a reasonable amount of time, in all likelihood, they do require more intensive intervention. Some students return to primary prevention and do not do well there, and therefore need to return to secondary intervention, return to tutoring again, because decisions have been incorrect about the fact that they are responders. In fact, they are not responders, and they emerge again as requiring secondary prevention. We want, to the extent possible, to avoid this ping-ponging back and forth between primary and secondary prevention with accurate, valid decision rules for determining who is and is not responding.

Slide 4: Another reason why it's important to quantify response using valid decision rules is that

research shows us that different methods and measures for designating response produce different decisions about who returns to general education verses who moves on to more intensive intervention. And in a related way, some methods and measures provide a more valid basis than others for making these decisions.

Slide 5: So let's talk for a few minutes about what some options are for distinguishing responders from non-responders. One popular method for designating response is to use a student's final status after tutoring ends. And often, final adequate status is defined in terms of a normative framework and a popular criterion for designating response is that the student perform above the 25th percentile, which is a standard score of 90. In some measurement systems, it is also possible to define final status at the end of tutoring in terms of a benchmark or a criterion that is associated with future success on an important outcome. Now a potential problem of using final status, post-tutoring status, is that it indexes final status but fails to reflect learning. So, it is possible that a student with decent final status may show worse growth than other students do not have adequate final status. Some people argue for a dual criterion that not only considers final status at the end of tutoring, but also the amount of growth that a student has manifested over the course of tutoring.

Slide 6: Now some people just use slope of improvement during tutoring and often this is defined as the weekly rate of improvement based on a progress monitoring measure so that we're using one progress monitoring probe per week or two probes every two weeks, and we usually have an assessment method, a progress monitoring method, that gives us a normative cut point, that tells us what an adequate rate of improvement is for that measure. Now in this situation, we have a good index of learning, but it doesn't tell us any information about whether the student's final post tutoring status is adequate, and some students with poor slope may in fact show good final status.

Slide 7: So, that brings us to the next slide where some people recommend focusing on a combination of final status and slope, and so we define response as adequate slope plus strong performance on the average of the last two progress monitoring scores. And these cut points for slope and for final status on the progress monitoring measure are usually set normatively. So, in terms of a representative sample of students at that grade level, and in this case, we index learning as well as final status and in most systems if either final learning or final status is met, then response is considered to be adequate.

Slide 8: Now, whatever you decide to use in terms of the measure and the method that you apply to that measure – that is, once you've picked a measure, are you going use final status, are you going to use slope of improvement, or are you going to consider a combination of slope and final

status – each method you select or you consider, very well may produce different decisions about which students are responsive and which students are not responsive and for each method the assessment tool also makes a difference.

Slide 9: So, in some different methods and measures will result in different groups of students being designated as responsive verses unresponsive. Some methods and measures will produce groups that are more valid for distinguishing students with and without severe learning deficits, unimportant, external criteria at the end of tutoring. Different methods and measures will also produce groups that are more valid than others for distinguishing which students will go on to develop serious difficulty in the later grades. And some methods and measures will produce groups that give us a more realistic prevalence of students who do have serious learning problems.

Slide 10: So let's look at a little bit of data. I'll briefly share with you two illustrative studies: one in reading and one in math. Here we see an overview of a study that Doug Fuchs and colleagues conducted where methods for designating response were contrasted in terms of how well they corresponded with later reading disability status. So 252 at-risk 1st graders received 10 weeks of small group tutoring. This tutoring was conducted in 1st grade, but then the students were followed to the end of 2nd grade. Now at the end of tutoring, the researchers designated responsive verses unresponsive students, and then we looked to see how the options for distinguishing responders from non-responders compared in terms of the prevalence of unresponsiveness, the accuracy of distinguishing reading disability from students who turned out not to have reading disability at the end of second grade, a year later after tutoring ended. And also, have a message for distinguishing response verses non-response compared in terms of the severity of students' learning problems at the end of second grade. So did the unresponsive students… how severe were their deficits compared to the students who had been designated as responsive?

Slide 11: So here we see a few different options that were considered in Doug Fuchs' 2008 paper. And so what you see if you look over to the last column is that we're contrasting 4 different methods on this slide. So there is final normalization, the one method for designating responsive verses unresponsive students involved taking a student at the end of tutoring in 1st grade and looking at the Woodcock Word ID measure and taking students who were above the 25th percentile on that measure and calling them responsive, and the students who were below the 25th percentile were called unresponsive. So we looked to see how well that method for designating response worked in terms of the percentage of students who were designated unresponsive. You see that a very small percentage of students were designated unresponsive

based on the Woodcock Word ID, only a little more than 2% of the general population. If we look at sensitivity and specificity with respect to end of 2nd grade reading disability status, we see that, that method for designating response at the end of 1st grade identified students who turned out later to be reading disabled, but, in terms of specificity, it designated many more students who actually turned out to be ok at the end of 2nd grade and erroneously designated those students as unresponsive. So, using the Woodcock Word ID measure in combination with a cut-point of the 25th percentile at the end of tutoring, we would have put many students into tertiary prevention when, in fact, those students looked fine at the end of 2nd grade. Now, if we look at CTOPP Site Word Efficiency, the comprehensive test of phonological processing site word efficiency, and again we're using a cut-point of above or below the 25th percentile to designate responsive versus unresponsive, we see that using site word efficiency we do much better in terms of a realistic prevalence for unresponsiveness - which is about 5% - and in terms of the designation of responsive and unresponsive at the end of 1st grade aligning with how students looked in terms of reading disability at the end of 2nd grade, we see that site word efficiency does better than Word ID because it is being pretty good at being sensitive to when students are in fact reading disabled at the 2nd grade. At the same time it does a pretty good job at being specific at who turns out to have a reading disability. That is, it is not giving us a large number of false alarms: students who are being designated unresponsive, probably put into tertiary prevention even though they would have done fine at the end of 2nd grade without that tertiary prevention. Now in the same way if you look at the numbers for word identification fluency slope and word identification slope plus final status - that is the dual discrepancy criterion – you see that both of those ways of indexing responsiveness, designating responsiveness, also do very nicely as the site word efficiency measure does, better than Woodcock word ID measure. So if you go to that paper, you will be able to look to see how these as well as many other measures for designating responsive verses unresponsive compare in terms of the prevalence of unresponsiveness, how well those measures and procedures for designating unresponsiveness correspond to reading disability status a year later after tutoring ends, and also that last column shows how severe, is an index of how severe the response of students' reading deficits at the end of 2nd grade were compared to the reading deficits of students who were designated as responsive at the end of 1st grade.

Slide 12: So here is an example of a math study also conducted at 1st grade. 127 at-risk 1st graders were randomly assigned to a control group in which they did not receive tutoring or to receive 16 weeks of small group tutoring. We progress monitored these students weekly using a curriculum based measurement computation measure and then we compared options for quantifying responsiveness to tutoring. So, options for designating responsive verses unresponsive students when tutoring ended.

Slide 13: So two broad conclusions about math options at 1st grade are these: for final status at the end of tutoring, achievement measures with a very thorough sampling of the skills that are taught at 1st grade level may produce sound decisions about who does and does not respond to intervention. That concerns what kind of measures do we want to use in math, at 1st grade? We want to use measures that comprehensively sample the kind of skills that are taught at 1st grade. Then, we have to apply a method to those measures for designating response and based on this study, final low achievement on 1st grade curriculum based measurement concepts and applications produced sound, valid decisions for distinguishing students who did and did not respond to the 1st grade tutoring as did a dual discrepancy. That is, slope plus final status on CBM computation.

Slide 14: So, let's take a closer look at progress monitoring with CBM computation to figure out method, gain more insight into how we might want to think about designating students as responsive verses unresponsive. So, in this same study, classroom teachers administered CBM computation each week for 27 weeks and those CBM computation tests were administered to at-risk as well as not at-risk students. Then, a year after tutoring ended, at the end of 2nd grade, we assorted the children into those who did and did not have a math disability. We defined MD – math disability – as performing below the 15th percentile across computation and word problems. Now, among students who had failed the 1st grade screen on CBM computation, a higher proportion of at-risk students without tutoring were in fact MD by the end of 2nd grade, as we might expect, if the tutoring program is effective. But the question here is: how well does CBM computation initial level – that is, the 1st grade screen – and the CBM computation slope of improvement – that is, those 27 progress monitoring tests that we collected over the course of 1st grade – how does screen and progress monitoring forecast MD verses not-MD at the end of 2nd grade?

Slide 15: So what you see here are students who all passed the screen. These are students who are, based on the screening, not at-risk for poor performance when they begin 1st grade. And you can see that we have a higher and lower growing group of students but all of these students are having very nice growth trajectories over the course of 1st grade. They are progressing well. So, if they pass the screen, the progress monitoring data do not really add very much information. You don't need to really progress monitor those students. If they pass the screen, then they do fine and essentially none of them end up, at the end of 2nd grade, with a math disability.

Slide 16: If you look at these blue lines what you see are students who failed the 1st grade screen. They were our at-risk students in this study. This is the group of students, these two lines, include the students who were chosen assigned to participate in the control group. And we have two different, distinct patterns of development on the progress monitoring data. For these students: none of them received tutoring. We have a small group of students – 16 students – who

actually, despite their low screen score at the beginning of 1st grade, developed nicely during 1st grade. That is in the blue solid line. But then we have a larger group of 46 students who failed the screen at the beginning of 1st grade and in fact developed inadequately during 1st grade and by the end of 2nd grade a very high proportion of students in that dotted line turn out to have a math disability.

Slide 17: Now I have two lines like that, these are the green lines, for the at-risk students who began the study, all those students failed the screen, all the green and all the blue, but these students in green, they actually received tutoring. And what we see is a nice strong, solid green line which indicates good development. Now these are the tutored students, and it's a very big group of students. 47 out of 61 are developing nicely in 1st grade. But the green dotted line, there is a small group of students who received tutoring, 14 out of 61, who have inadequate progress monitoring data. Their trajectory of development does not look good. And at the end of 2nd grade, a large number of those 14 students who progressed poorly despite tutoring, end up having a math disability at the end of 2nd grade.

Slide 18: So there are some take home points from these 6 trajectories of development over the course of 1st grade. The first take home message is that students who pass the initial CBM computation screen do not turn out to have a math disability at the end of 2nd grade. Students who pass the screen, if it's a good cut point, therefore, do not require progress monitoring. We have enough information just knowing that they passed the screen. And their math development, at least through the end of 2nd grade, is strong. The problem is that when students fail the screen, the screening score alone is not accurate for predicting math disability status at the end of 2nd grade.

Slide 17: So if we look back at our slide, we have four groups, all the blue and all the green lines. They all failed the screen, and their screening data alone does not tell us how well they're going to develop over the course of 1st grade. Instead we need the progress monitoring data, those 27 data points, to be able to model their 1st grade math development and their 1st grade math development then corresponds well with how they look in terms of MD status at the end of 2nd grade.

Slide 18: So students below the screening cut-point require progress monitoring. Some of those students will progress nicely, and others will not, and this is true whether or not they receive tutoring. So CBM computation progress monitoring is useful for distinguishing which students will and will not have severe forms of math difficulty at the end of 2nd grade.

Slide 19: So across reading and math, to distinguish responders from non-responders, the method – are you using final status, are you using slope of improvement – the method makes a

difference in how sound your decisions about responsiveness verses unresponsiveness will be. Assessment tool also makes a difference. So we need to select a method and an assessment tool carefully to understand how accurate that method and tool will be for identifying students who in the long term can be expected to fare well in general education. We want those students to accurately be designated as responsive to secondary prevention. And we need to select a method of assessment tool that also accurately identifies for us students who will not fare well unless they receive more intensive tertiary prevention.

Slide 20: Now one source that's really great for giving us information about what tools to consider is to go to the National Center on RTI's progress monitoring tools chart. So the NCRTI has 3 different tools charts. One of those tools charts is on progress monitoring. And from going to that website, we can learn information about what method various tools recommend for quantifying response when you're using their tool, and how good the research data are to support their recommendation for the method that the tools employs for designating responsive verses unresponsiveness.

Slide 21: If we go for example to the technical review committee for the National Center on RTI requires vendors to specify standards for minimal acceptable growth when we're progress-monitoring students. This serves as one potential method for designated responsive verses unresponsive students. So the question that is being asked is: "Is minimum acceptable growth well specified for that tool?" And the growth standards can be norm-referenced or they can be criterion-referenced. So, let's look at a few examples from the progress monitoring tools chart for the National Center on RTI.

Slide 22: Here we see information on a tool: curriculum based measurement letter-sound fluency in kindergarten. That tool has a recommendation for how users can determine response. If you look at the 3rd and 4th column on this slide, you see that to determine response the vendor for that tool based on research data that they have, is recommending that post-tutoring status should be at least 30 letter sounds correct in a minute, and the slope of improvement over the course of tutoring should be improvement of at least 1 letter sound per week, that slope of improvement of 1.0.

Slide 23: So the technical review committee looked at that tool and looked at the recommendations that the tool provides for designating responsiveness, and the technical review committee found that recommendation to be convincing based on the data that the vendor provided to support that recommendation.

Slide 24: Here is another example. We have a STEEP measure, and here we have recommended weekly rates of improvement at grades 1 through 5. So the measure STEEP, the vendors of that measure, submitted the tool to the technical review committee. They provided these

recommendations for determining adequate weekly rates of improvement for designating responsive verses unresponsive students. And they provided their research data to support their recommendations for weekly improvement.

Slide 25: And here again the technical review committee found those recommendations for the STEEP tool to be convincing.

Slide 26: And here we see a 3rd example: the STAR early literacy measures. Here we have growth expectations like slopes of improvement, and here the vendor is providing slopes for students who are below the 20th percentile and for students who are at the 50th percentile. The STAR people provided research documentation to support these recommendations.

Slide 27: And the technical review committee, again, rated these recommendations and the data supporting the recommendations to be convincing.

Slide 28: So if you go to the tools chart on progress monitoring you can look at a lot of different kinds of information, including rates of improvement, which is the 4th column toward the end of the chart, rates of improvement specified. And if you see a solid dot, it tells you that the technical review committee found that tools recommendation to be convincing, but if you see a half full or an open dot, then the technical review committee did not find the recommendations to be convincing. And you can look down any of the columns for different kinds of information that the technical review committee judges, to see how well the tool fared. So here you see AIMSweb…

Slide 29: …and you can look down the chart to see curriculum based measurement in the area of reading, DIEBELS measures are on the chart…

Slide 30: …a measure called easyCBM, mClass…

Slide 31: …Monitoring Basic Skills Progress, some Orchard software products, a product by Scholastic, and then the STAR and STEEP measures, were all evaluated by the technical review committee and you can get lots of good information from the tools chart, including methods for designating responsive verses unresponsive students. So that gives you a little bit of information about how we think of responsiveness verses unresponsiveness. At the end of secondary prevention, at the end of tutoring, schools are faced with making a decision: Who has been responsive to this tutoring? Who goes back to general education? – with the assumption that they will do well there. And which group of students has responded to tutoring inadequately, such that we designate those students as unresponsiveness and because of that we move those students into a more intensive form of intervention. So we've talked today about how to think, from a

measurement perspective, about what measures to use and what methods to apply to the data that are derived from those measures for designating responsive verses unresponsive students.

Slide 32: If you're interested in getting more information about the National Center on RTI, you see the link, the website on the slide and then you can navigate to the progress monitoring tools chart that I talked a little bit about on the center's website and you see also the link on this slide.

Slide 33: And if you have any questions you can email them to RTIWebinars@air.org.

Slide 34: So thank you!