

# Clinical Significance: An Exploration into a Methodology for Tier 2 Placement within the Response to Intervention Framework

---

Rachel Lacks  
Jiawei Li  
Sandee McClowry  
New York University, United States

## Abstract

*Based on a Response to Intervention framework, this study examines whether a clinical significance methodology is useful for placing children in Tier 2 interventions for school-based social-emotional learning. Data from INSIGHTS Into Children's Temperament, a Tier-1 intervention, was analyzed to assess practical considerations and limitations inherent in the clinical significance methodology. Student academic and behavioral outcomes were analyzed using the Reliable Change Index (RCI) to assess recovery and reliability. Based on findings that demonstrate that the number of students appropriate for Tier 2 is consistent with prior research, the clinical significance methodology is discussed as a feasible and easy-to-execute way for school personnel to determine student responsiveness to interventions.*

**Keywords:** Response to intervention, social emotional learning intervention, clinical significance, disruptive behaviors, academic achievement

## Introduction

Children who are self-regulated are engaged in the learning process and exhibit fewer disruptive behaviors in the classroom (Diamond & Lee, 2011). To enhance children's self-regulation, schools are increasingly implementing social-emotional learning (SEL) programs (Durlak, Weissberg, Dymnicki,

Taylor, & Schellinger, 2011). Although there are a number of evidence-based SEL intervention programs, most are focused on the classroom level (Durlak et al., 2011; Korpershoek, Harms, de Boer, van Kuijk, & Doolaard, 2016). As a result, they are limited in their ability to effectively meet the individual needs of students who require more intensive services (Kendziora & Yoder, 2016). Ten to 15% of students in general education classrooms fail to demonstrate adequate improvements post-intervention, and require more specialized instruction to better suit their needs (Fox, Carta, Strain, Dunlap, & Hemmeter, 2010). Multi-tiered models of prevention, alternatively, aim to reach all students with specialized levels of instruction (Bruhn, Lane, & Hirsch, 2014).

Response to Intervention (RtI) is a multi-tier framework for the early identification and support of students with varying learning needs. One of its aims is to differentiate students who are struggling because of low-quality instruction versus those with actual disability (Fuchs, Fuchs, & Compton, 2004). Although the majority of the RtI literature and related discussion have focused on academic outcomes, specifically learning disorders (LDs) (Fuchs & Deshler, 2007; Fuchs, Mock, Morgan, & Young, 2003), research has emerged to support the application of this tiered model to social and behavioral goals (Cheney, Flower, & Templeton, 2008; Fairbanks, Sugai, Guardino, & Lathrop, 2007; McIntosh, Campbell, Carter, & Dickey, 2009).

The RtI model is intended to monitor student progress and provide appropriate accommodations through the provision of multiple tiers of intervention (McIntosh et al., 2009). In Tier 1, students receive general education instruction. Those who do not respond adequately are deemed “non-responsive” and are placed into subsequent tiers to receive more specialized treatment (Fuchs & Fuchs, 2006). In an SEL context, Tier 1 interventions provide core curriculum programming for all students in a classroom. Tier 2 is provided in addition to Tier 1 for students at higher risk for social emotional delays who require secondary interventions beyond the core curriculum. Tier 3 provides tertiary treatments that are highly individualized and more intensive for children with persistent challenges (Fox et al., 2010).

An RtI framework supports a movement away from looking for within-child deficits to an examination of the broader school context and structural problems that maybe preventing optimal student achievement (Harris-Murri, King, & Rostenberg, 2006). Moreover, RtI models seek to ensure that students receive evidence-based interventions to prevent unnecessary referrals to special education. In many schools, however, the response to a student’s behavioral problems is often compounded by other factors. When teachers feel inadequately equipped to handle disruptive behaviors in the classroom, a referral to special education is often the remedy (Skiba et al., 2006).

However, the effect is not being seen across all students equally. Students of color are often largely overrepresented. Indeed the Nondiscriminatory Administration of School Discipline Policy of 2014 issued by the

Departments of Education and Justice brought to light the racial disparities that are evident in special education referrals and disciplinary actions (US Department of Justice & US Department of Education, 2014). Data regarding office discipline referrals in 364 elementary and middle schools nationwide revealed that students from African American families are 2.19 times as likely to be referred to the office for disruptive behavior as are their white counterparts in elementary school (Skiba et al., 2011). In middle school, African American students are 3.78 times more likely to be referred. Other factors in addition to race are associated with special education and discipline referrals. Students from low socio-economic families or who have academic problems are disproportionately punitively disciplined in schools or placed into special education. Children with more than one of these demographics are at even greater risk. A direct link has been heavily researched between these exclusionary discipline tactics and entrance into the juvenile justice system; the link has been termed the “school-to-prison pipeline” (Fenning & Rose, 2007).

Tiered social emotional interventions are needed to meet the varying needs of students, especially those from diverse backgrounds. Likewise, practical strategic methods are necessary for identifying children requiring different levels of SEL programming. This paper seeks to explore an applied methodology for designating students to tiers of SEL interventions that is easy for teachers to use.

## **Existing Methodological Challenges in the RtI Field**

Fuchs, Fuchs, and their colleagues have been candid in their acknowledgement that the RtI field has a long way to go. The RtI framework is a promising start for the identification of students at risk for learning, behavioral, or other related problems, but it is far from perfect. The greatest methodological challenge in the RtI field is identifying the most effective method to use for designating students to tiers of intervention (Fuchs & Fuchs 2006). In a related manner, there is a lack of clarity around the R in RtI – what constitutes responsiveness (Speece & Walker, 2007)? The current use of multiple methods for evaluating responsiveness is problematic in that it results in disparate categorizations of students and differing prevalence rates of students identified as non-responsive (Fuchs et al., 2004).

The RtI literature describes several distinct methodologies to categorize students as responsive or non-responsive. These methods are frequently applied when identifying children with LDs, rather than for use in SEL programming. Two such methodologies use post-treatment status to evaluate responsiveness. The “final benchmark” (Fuchs & Dechsler, 2007) method compares raw scores on a measure against a criterion-referenced benchmark associated with future success; students whose scores exceed a particular threshold are designated as responsive. For example, Torgesen and colleagues (2001) used the “normalized” post-treatment standard score;

those who were at or above the 25th percentile at post-test were responsive while those below were deemed non-responsive. Similarly, Good, Simmons and Kame'enui (2001) looked at performance at the end of the intervention. Alternatively, Vellutino and others (1996) evaluated student performance several times throughout a multi-year program using hierarchical linear modeling to calculate the slope of improvement for each child on multiple measures. The slopes were then rank ordered, and by way of the "median split method" (Fuchs & Dechsler, 2007), the median was the point used for the cutoff between responsive and non-responsive.

Another method developed by Fuchs and Fuchs (1998) and utilized by Speece and Case (2001) is termed the "dual discrepancy" method (Fuchs & Dechsler, 2007). Students are assessed in two ways: slope of improvement during intervention and performance level at the end of the intervention. Students designated as non-responsive scored more than one standard deviation below peers on both slope and performance level at the end of intervention. Lastly, a method used by Fuchs et al. (2004) based responsiveness on students' slope of academic improvement. Students were measured over a period of time and their slopes were computed. Students who scored above a normative cut off point were responsive, and those below were non-responsive.

Fuchs and Deshler (2007, p.134) conclude that there are "a handful of possibilities but no consensus about which are preferable." Consensus may not be a feasible goal, however, given the different purposes an RtI framework is used to accommodate. The aforementioned methodologies were applied to educational interventions intended to produce academic improvement, usually reading. They may not prove useful in assessing individual-level change to SEL-focused interventions. The methodologies that are currently in practice also have a number of drawbacks that make them difficult to utilize in school settings. Many of these methods are highly time- and resource-heavy because they require the collection of data at multiple time points for analysis (Fuchs & Fuchs, 1998; Speece & Case, 2001). Relatedly, the methodologies often require knowledge of sophisticated statistical analyses that are outside the scope of a typical teacher's training. In low-income schools in particular, in which teachers are over-worked and under-resourced (Selwyn, 2007), the expectation to use complex methodologies for student tier designation is unrealistic.

Another method, thus far untested in the RtI context, is proposed as a feasible way for teachers to assess students' post-intervention change. The authors borrow a method from clinical psychology and the broader medical fields to assess individual response to intervention. The clinical significance methodology, as it will be referred to as, represents a movement away from traditional methods of evaluating change by examining data on an individual level. It allows post-hoc exploration into why some individuals did or did not improve post-intervention and may be helpful in addressing factors that

mediate and moderate improvement (Gerdes, Haack, & Schneider, 2012). The clinical significance methodology is not resource-intensive and lends itself to school adaptation past the pilot phase of an intervention.

## **Clinical Significance Methodology**

When applying the clinical significance methodology, clinically significant change must meet two essential criteria: recovery and reliability (Jacobson & Truax, 1991). Individuals are recovered if they initially score in a clinical, or dysfunctional, range on a particular measure, and then score in a normative, functional range during a post-treatment assessment of the same outcome. Clinically meaningful change is concluded when a change has been made from the dysfunctional to the functional range when this change is reliable, or unlikely to be the product of error. Testing for reliability safeguards against the possibility of an individual's post-test score crossing the cutoff point but not to an extent that is statistically significant.

This methodology groups individuals into one of several categories: a) if the individual shows statistically reliable change post-treatment but remains in the dysfunctional range, the individual is "improved but not recovered," b) if the individual is in the functional range post-treatment, but the magnitude of change is not statistically reliable, then the methodology is not able to determine if this change was clinically significant, a random fluctuation, or a measurement error, and c) if the individual demonstrates change that is both statistically reliable and puts the individual within a functional range, the individual is "recovered" (Jacobson, Roberts, Berns, & McGlinchey, 1999). The clinical significance methodology allows a student to be compared to him or herself at an earlier time point, thus reducing bias by not comparing diverse groups of students against each other.

The terminology in the clinical significance literature requires some adaptation for a school context. Jacobson and Truax (1991) and others use the term "functional" to refer to clients who score in the normative range on a particular measure and "dysfunctional" for those who score outside of the normative range. In the current paper, we make the case that those individuals who score as functional on behavioral and academic outcomes based on the clinical significance methodology are "responsive" to intervention while those identified as dysfunctional are "non-responsive."

## **Current Study**

The purpose of this study was to explore whether calculating the clinical significance of student outcomes from an efficacious Tier 1 intervention could inform the creation of multiple tiers as recommended by RtI. We focus specifically on identifying students who need Tier 2 services because 10-15% of the student population are hypothesized to need these secondary supports in an RtI context (Bruhn et al., 2014). To achieve this aim, the data from a randomized clinical trial of a Tier 1 intervention, *INSIGHTS Into Children's Temperament (INSIGHTS)*, was examined.

*INSIGHTS* is a comprehensive school-based universal intervention that aims to foster social and emotional learning competencies in children in order to promote their self-regulation and academic achievement (McClowry, 2014). *INSIGHTS* includes classroom, teacher, and parent programs intended to support children's ability to self-regulate by improving their attentional and behavioral capacities.

In a group randomized trial, *INSIGHTS* was compared to a supplemental reading program in enhancing the academic learning context (O'Connor, Cappella, McCormick, & McClowry, 2014). Two level individual growth-modeling demonstrated that children in *INSIGHTS* had a 1.23-point increase in math scores ( $ES = .31$ ) and a 3.15-point increase in reading scores ( $ES = .55$ ) on the Woodcock-Johnson III Tests of Achievement. In addition, children in *INSIGHTS* exhibited reductions in behavior problems over time ( $ES = .54$ ) compared with children in the supplemental program who showed increased disruptive behavior over this same time period.

Interventions such as *INSIGHTS* demonstrate efficacy in improving academic and behavioral outcomes on the group level, but do not identify which students require additional services. Making use of RtI can help to ensure that individual students in general education classrooms receive a level of intervention that is matched to his or her level of need (Fox et al., 2010).

## Method

### Participants and Setting

The randomized clinical trial of *INSIGHTS* was conducted in 22 under-resourced urban schools (see O'Connor et al., 2014 for a full study description). The participants included 435 children and their parents or caretakers as well as 122 teachers from their kindergarten and first grade classrooms. Eleven of the schools were randomly assigned to *INSIGHTS*. The remaining schools participated in a supplemental reading program, which will be referred to as the control group.

At baseline, the children ranged in age from 4 to 7 years ( $M=5.38$  years,  $SD=0.61$ ). Fifty-two percent of the children were boys, and the remaining 48% were girls. Eighty-seven percent of the children qualified for free or reduced-price lunch programs. Approximately 75% of the children were Black, non-Hispanic; 16% were Hispanic, non-Black; and the remaining children were biracial.

### Data Collection

Researchers and associated staff received training on all measures that were administered prior to the five data collection time points. Data collectors had no knowledge of school study condition. For this analysis in the current study, we use data collected at baseline in the winter of the kindergarten year. Then the intervention was conducted in the Spring during kindergarten

and repeated during the following Fall when the children were in first grade. The post-intervention data was collected in the winter of the following year, immediately following the first grade intervention.

## Measures

Multiple measures were administered in the third randomized clinical trial of *INSIGHTS* (O'Connor et al., 2014). For the purposes of this paper, outcomes in student's behavioral problems and academic achievement were examined. These areas were chosen because *INSIGHTS* is designed to increase students' abilities to self-regulate which acts to improve academic and behavioral outcomes. Making use of both behavioral and academic data is useful in providing a more comprehensive assessment of the student because behavior and academic performance mutually influence each other (Lane, Menzies, Kalberg, & Oakes, 2012; Lane & Wehby, 2002).

**Child behavior problems.** Disruptive behavior problems were measured with the 36-item Intensity Scale of the Sutter-Eyberg Student Behavior Inventory (SESBI) (Eyberg & Pincus, 1999). On a frequency scale ranging from 1-7 (1=never, 3=seldom, 5=sometimes, 7=always), teachers report the frequency that a student engages in a variety of disruptive behaviors, such as "acts defiant when told to do something," and "verbally fights with other students." An aggregate score was calculated by adding the number of individual items to create a total frequency score. Cronbach's alpha was .96 for the Intensity Scale.

**Child academic achievement.** Reading and math achievement were assessed using the Letter-Word Identification (LWID) and Applied Problems (AP) subtests of the Woodcock-Johnson III (WJ-III) (Woodcock, McGrew, & Mather, 2001). The Letter-Word Identification subtest assesses letter naming and word decoding skills by asking children to identify a series of letters and words. The Applied Problems subtest assesses children's simple counting skills and the ability to analyze and solve mathematical word problems. Their average reliabilities in this study were .88 and .84 respectively.

## Analytic Approach

A clinical significance score was calculated to determine individual student's progress from baseline to immediate post-intervention in first grade. Students who did not have both baseline and post-intervention data (i.e. students who received the intervention only in first grade or who had missing data for other reasons) were not included in analyses.

A Reliable Change Index (RCI) was calculated for each individual student (Christensen & Mendoza, 1986; Jacobson, Follette, & Revenstorf, 1984). The RCI assesses the reliability of change an individual demonstrates over time in order to demonstrate that the change was not merely a random fluctuation or measurement error (Jacobson & Truax, 1991). Using the Reliable & Clinical Changer Generator for Windows, version 4 (Deville, 2004), the RCI was calculated by dividing the pre-post-intervention score difference by the standard error of the difference between these two scores for each measure.

An RCI > 1.96 indicates reliable change ( $p < 0.05$ ). Recovery was defined by clinical cutoff scores based on the midpoint between the mean of the functional population and the mean of the dysfunctional population. The midpoint calculation takes into account the unequal variances of functional and dysfunctional populations (Jacobson & Truax, 1991).

Students were then classified into one of four reliable categories post-intervention: Category 1) functional at baseline and remained functional post-intervention; Category 2) functional at baseline and reliably moved to non-functional range at post-intervention; Category 3) non-functional at baseline and remained non-functional at post-intervention; Category 4) non-functional at baseline and reliably recovered to functional range at post-intervention (clinically significant change). Participants who did not change reliably, but who did cross the threshold from either functional to non-functional or from non-functional to functional were in Category 5.

## Results

The frequencies and percentages of student functionality both pre- and post-intervention are reported for student behavior problems, math achievement, and reading achievement, and are shown in Table 1. Figure 1 (next page) provides a visual representation of Tier 2 designation based on student functional status and responsiveness to intervention.

The number of students who were in *INSIGHTS* and reliably completed the intervention in the functional range (Categories 1 and 4) was high across all three variables. These children were deemed responsive to the intervention. They included 80% of the children whose behavior was functional based on the SESBI. On the WJ III AP, 83% scored in the functional range in their counting and skill in solving mathematical word problems. Regarding the WJ III LWID, 89.3% scored in the functional range on their ability to identify letters and read words of increasing difficulty at post-intervention.

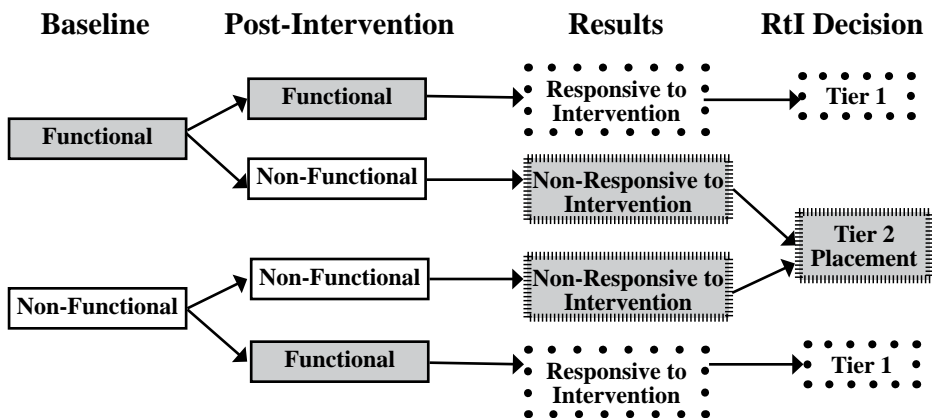
Table 1

### *Frequency of Clinical Significance of Child Behavior Problems and Academic Achievement*

Variable	Category 1 Functional to Functional <i>n</i> (%)	Category 2 Functional to Non- Functional <i>n</i> (%)	Category 3 Non-Functional to Non- Functional <i>n</i> (%)	Category 4 Non-Functional to Functional <i>n</i> (%)	Category 5 Non-Reliable Change <i>n</i> (%)
Child Behavior Problems	79 (74.5)	13 (12.3)	8 (7.6)	6 (5.7)	0 (0.0)
Applied Problems	97 (80.2)	2 (1.7)	8 (6.6)	3 (2.5)	11 (9.1)
Letter-Word Identification	106 (87.6)	2 (1.7)	7 (5.8)	2 (1.7)	4 (3.3)

Note: Child Behavior Problems ( $n=106$ ), Applied Problems and Letter-Word Identification ( $n=121$ )



**Figure 1.** Student Responsiveness to Intervention and Tier Placement Decision Model

Note: Students who demonstrated non-reliable change was not included in Figure 1.

Students who reliably scored in the non-functional range (Categories 2 and 3) were non-responsive to intervention and qualified for a referral to a Tier 2 level SEL program. Approximately 20% of children ended the intervention in the non-functional range in terms of behavioral problems on the SESBI. Because social emotional behavior is interrelated with academic skills (Lane et al., 2012), non-functional scores on the WJ III were also considered. Eight percent of the children were reliably in non-functional range on Applied Problems post-intervention and 7% were reliably non-functional on Letter-Word Identification.

It should be noted that we identify an area of ambiguity inherent to the clinical significance method that will be further addressed in the Discussion section: not all students demonstrated reliable change at post-test. Category 5 includes some students whose post-intervention score was in the non-functional range, but who demonstrated a degree of change from baseline to post-intervention that was not statistically reliable.

## Discussion

To our knowledge, this is the first study that assesses clinical significance (both reliable change and recovery) as a methodology for creating tiers in school-based SEL interventions. Categories of responsiveness are useful if they have accurate ability to predict students who are “on track” to meet designated benchmarks (Silberglitt & Hintze, 2005). The ultimate purpose in determining cut off points is to identify children who are likely to exhibit disruptive behavior in the future that will preclude them from participating effectively within the classroom setting and may manifest in a host of negative later-life outcomes (Becker & Luthar, 2002; Fuchs et al., 2004).

Despite the fact that several methods have been utilized to assess student

responsiveness to intervention, the lack of consensus amongst practitioners on a preferred method speaks to the limitations inherent to the existing methods (Fuchs & Dechsler, 2007; Fuchs & Fuchs, 2006). The gap also highlights the limitations in currently used methods in their applicability in schools.

One of the greatest strengths of the clinical significance methodology is its simplicity: the method requires assessment at only two time-points, rudimentary statistical knowledge and analysis, and an unspecified sample size because data is examined on an individual level. In contrast, currently existent methodologies may not be feasible for use in schools after an intervention because of their need for strict adherence to multiple time-consuming guidelines, their requirements of large sample sizes, as well as an excessive amount of additional work for already over-worked teachers. These findings all point to the need for an efficient, simple-to-execute, and most importantly, realistic to sustain, method of determining student eligibility for specific tier services, which clinical significance lends itself to well.

Results of this study have implications for utilizing this particular method in school-based intervention work. The clinical significance methodology in this study produced results that are consistent with previous literature's indication of the proportion of students that qualify for Tier 2 services (Fox et al., 2010). Based on the results, approximately 20% of students would be referred to a Tier 2 SEL program. This data is congruent with, albeit slightly higher than the 10-15% of children that research indicates are not optimally responsive to universal interventions (Bruhn et al., 2014; Fox et al., 2010).

The high-risk nature of the sample also may account for the slightly elevated frequencies of students that were not optimally responsive to intervention; 10.7% WJ LWID, 15.7% WJ AP, 19.9% SESBI. The study is unique in the demographic make-up of its sample: urban, low-income, minority students. This high-risk population is likely a large referral source to RtI, and yet they are largely understudied in the RtI field (Graham & Perin, 2007; Reynolds & Shaywitz, 2009).

Relatedly, this examination of the clinical significance methodology for tier designation in this study makes an important contribution to the RtI field because it includes a predominantly black and Hispanic sample of urban students from low socioeconomic backgrounds. This population is overrepresented in school disciplinary and special education referrals compared to their white counterparts (Skiba et al., 2011). As schools face pressures to have students meet academic benchmarks in order to receive funding and to remain open (Bassok, Latham, & Rorem, 2016; Skiba et al., 2006), there has been greater use of exclusionary discipline practices that populate the school-to-prison pipeline with students of color (Fenning & Rose, 2007). RtI models seek to ensure that students receive intensive evidence-based intervention before a referral to special education is made as a means of combatting the inequality seen in the racial make-up of students with special education and discipline referrals (Cramer, 2015).

Despite the potential for this methodology to prove immensely useful in evaluating and creating tiers within SEL interventions, there are several limitations to consider when using the clinical significance strategy. First, of the utmost importance, is to recognize the drawback inherent to designating a student as having achieved clinically significant change: the student must have begun in a non-functional range at baseline. A majority of the students in regular education classrooms begin at a functional level, as opposed to students in special needs classes, therefore precluding them from making clinically significant change by definition of being unable to recover. If educators are interested in gauging responsiveness in the context of RTI as opposed to post-intervention functionality, we advise that practitioners focus solely on students who were in the non-functional range on the outcome at baseline to investigate this group's responsiveness to intervention in employing the clinical significance definition.

When classifying students at the end of the intervention in terms of responsiveness, it is important to bear in mind that students who did not demonstrate a clinically reliable degree of change cannot be left unclassified. Despite the ambiguity in scoring between the beginning and end of the intervention (i.e. Category 5, the score only just surpasses the threshold of responsiveness in either direction by a non-reliable amount), these students must ultimately be classified as either responsive or non-responsive based on their final score. In a research sense, these students may be left out of analyses, but in a practice sense, these students may not be simply left out of the intervention.

Acknowledging the questionability of the confidence with which we can declare their final score to be accurate, practitioners must make a decision in terms of classifying these students as responsive or non-responsive to intervention. This leaves room for subjective educator input, and can potentially result in a student receiving specialized services he does not necessarily need or, worse, a student in need of specialized services not receiving them. Ideally, practitioners want to provide the appropriate level of intervention to each student, but the vagueness in tier designation presents the threat of either: a) students not getting the more intensive dosage of intervention they need because their score slightly exceeded the threshold into the responsiveness range, or b) students using costly, individualized resources in a higher level tier than they need for optimal responsiveness because their score just missed the designation of responsiveness.

Missing data is another limitation that is encountered when using this method; clinical significance focuses on individuals, not group means. Missing data, therefore, cannot be imputed. This constraint reduces sample sizes that can be used for clinical significance studies when there is missing data. There also exists the strong possibility that individuals' scores on various assessments will not unequivocally designate them as responsive or not; an individual may score in the functional range on a measure of word identification, but may score in the non-functional range on a measure of behavior problems. We therefore strongly encourage practitioners to be

mindful of the criteria they set for of Tier 2 and 3 services (Ogles, Lambert, & Sawyer, 1995).

Effective and complete implementation of school-based interventions requires sustaining the intervention after the initial implementation by researchers (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). While this may sound obvious, SEL programs are often not implemented in a way that fits in with the day-to-day structure of schools, thereby limiting their long-term sustainability and effectiveness (Jones & Bouffard, 2012). The feasibility of adapting interventions into the school culture is necessary for the success of the intervention. It is essential to consider the generalizability from research to practice, as well as the real-life administrative and logistical issues that may arise during the implementation phase that were absent during a research study (Backer, Liberman, & Kuehnel, 1986; Schoenwald & Hoagwood, 2001).

Moving forward, it is imperative to test the long-term utility of the clinical significance method. In order to be a valid form of tier designation, the method should withstand the test of four diagnostic accuracy statistics, as outlined by Silberglitt and Hintze (2005): A) sensitivity: of all students who display future disruptive behavior problems, what percentage of them were accurately predicted to do so by earlier Tier 2 placement? B) specificity: of all students who do not display future disruptive behavior problems, what percentage of them were accurately predicted to do so by earlier Tier 1 placement? C) positive predictive power: of all students who were predicted to display future disruptive behavior problems by earlier Tier 2 placement, what percentage of them actually displayed future behavior problems? D) negative predictive power: of all the students who were predicted to not display future disruptive behavior problems by earlier Tier 1 placement, what percentage of them actually do not display later disruptive behavior problems?

This investigation highlights the need for future work to be done. There is still nowhere near a consensus on what operationalization of “non-responsive” should be used in intervention work. Collaborative efforts on the part of researchers, policymakers, and educators should aim to select valid and reliable criteria that can be employed in school settings. RtI has been touted as an alternate approach that could help to reduce the discrepancy in school discipline referrals between white students and students of color. The field of RtI is burgeoning and holds much promise; used in tandem with the clinical significance methodology, there is potential for a useful framework for informing tiered interventions. ■

## References

- Backer, T. E., Liberman, R. P., & Kuehnel, T. G. (1986). Dissemination and adoption of innovative psychosocial interventions. *Journal of Consulting and Clinical Psychology, 54*, 111–118.
- Bassok, D., Latham, S., & Rorem, A. (2016). Is Kindergarten the New First Grade?. *AERA Open, 2*(1).
- Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting achievement outcomes among disadvantaged students: Closing the achievement gap. *Educational psychologist, 37*(4), 197-214.
- Bruhn, A. L., Lane, K. L., & Hirsch, S. E. (2014). A review of tier 2 interventions conducted within multitiered models of behavioral prevention. *Journal of Emotional and Behavioral Disorders, 22*(3), 171-189.
- Cheney, D., Flower, A., & Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *The Journal of Special Education, 42*(2), 108-126.
- Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305-308.
- Cramer, L. (2015). Inequities of Intervention among Culturally and Linguistically Diverse Students. *Penn GSE Perspectives on Urban Education, 12*(1), n1.
- Devilley, G. J. (2004). The Reliable Change Generator for Windows version 2.0 (Version 2.0): The Centre for Neuropsychology, Swinburne University, Australia.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*(6045), 959-964.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development, 82*(1), 405-432.
- Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Behavior Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M. (2007). Response to intervention: An evaluation of a classroom system of behavior support for second grade students. *Exceptional Children, 73*, 288-310.
- Fenning, P., & Rose, J. (2007). Overrepresentation of African American students in exclusionary discipline the role of school policy. *Urban Education, 42*(6), 536-559.
- Fixsen, D. L., Naoom, S. F., Blase, K. B., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: National Implementation Research Network, Louis de la Parte Florida Mental Health Institute, USF.
- Fox, L., Carta, J., Strain, P. S., Dunlap, G., & Hemmeter, M. L. (2010). Response to intervention and the pyramid model. *Infants & Young Children, 23*(1), 3-13.
- Fuchs, D., & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice, 22*(2), 129-136.

- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93-99.
- Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice*.
- Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly*, 27(4), 216-227.
- Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, 18(3), 157-171.
- Gerdes, A. C., Haack, L. M., & Schneider, B. W. (2012). Importance of clinically meaningful change: Parental functioning in families of children with ADHD: Evidence for Behavioral Parent Training. *Journal of Attention Disorders*, 16, 147-156.
- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision making utility of a continuum of fluency based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288.
- Graham, S., & Perin, D. (2007). *Writing Next-Effective strategies to improve writing of adolescents in middle and high schools*. A report to Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education
- Harris-Murri, N., King, K., & Rostenberg, D. (2006). Reducing disproportionate minority representation in special education programs for students with emotional disturbances: Toward a culturally responsive response to intervention model. *Education and Treatment of Children*, 29(4), 779-799.
- Jacobson, N. S., Follette, W. C., & Revenstorff, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, 67, 300-307.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59(1), 12.
- Jones, S. M., & Bouffard, S. M. (2012). Social and Emotional Learning in Schools: From Programs to Strategies. *Social Policy Report*, 26(4). Society for Research in Child Development.
- Kendziora, K., & Yoder, N. (2016). When districts support and integrate social and emotional learning (SEL): Findings from an ongoing evaluation of districtwide implementation of SEL. Washington, DC: American Institutes for Research. *The WERA Educational Journal*, 31.
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management

- programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, 86(3), 643-680.
- Lane, K. L., Menzies, M., Kalberg, J. R., & Oakes, W. P. (2012). A comprehensive, integrated three-tier model to meet students' academic, behavioral, and social needs. In K. Harris, T. Urdan, & S. Graham (Eds.), *American Psychological Association Educational Psychology Handbook, Vol. 3* (pp. 551 – 581). Washington, DC: American Psychological Association.
- Lane, K. L., & Wehby, J. (2002). Addressing antisocial behavior in the schools: A call for action. *Academic Exchange Quarterly*, 6, 4–9.
- McClowry, S. G. (2014). *Temperament-based elementary classroom management*. Lanham, MD: Rowman & Littlefield.
- McIntosh, K., Campbell, A. L., Carter, D. R., & Dickey, C. R. (2009). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions*, 11(2), 82-93.
- O'Connor, E. E., Cappella, E., McCormick, M. P., & McClowry, S. G. (2014). An examination of the efficacy of INSIGHTS in enhancing the academic and behavioral development of children in early grades. *Journal of Educational Psychology*, 106(4), 1156.
- Ogles, B. M., Lambert, M. J., & Sawyer, J. D. (1995). Clinical significance of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Journal of Consulting and Clinical Psychology*, 63(2), 321.
- Reynolds, C. R., & Shaywitz, S. E. (2009). Response to Intervention: Ready or not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly*, 24(2), 130.
- Schoenwald, S. K. & Hoagwood, K. (2001). Effectiveness, transportability and dissemination of interventions: What matters when? *Psychiatric Services*, 52, 1190–1197.
- Selwyn, D. (2007). Highly quantified teachers NCLB and teacher education. *Journal of Teacher Education*, 58(2), 124-137.
- Silberglitt, B., & Hintze, J. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment*, 23(4), 304-325.
- Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School Psychology Review*, 40(1), 85.
- Skiba, R., Simmons, A., Ritter, S., Kohler, K., Henderson, M., & Wu, T. (2006). The context of minority disproportionality: Practitioner perspectives on special education referral. *Teachers College Record*, 108(7), 1424.
- Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*, 93(4), 735-749.
- Speece, D. L., & Walker, C. Y. (2007). What are the issues in Response to Intervention research? In: Haager D, Klingner J, Vaughan S, (Eds). *Evidence-based reading practices for response to intervention*. Baltimore: Paul H. Brookes; pp. 287–301.

- Torgesen, J. K., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and longterm outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*(1), 33–58.
- US Department of Justice & US Department of Education (2014). Dear Colleague Letter: Nondiscriminatory Administration of School Discipline Washington, DC: U.S.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S., Chen, R., Pratt, A., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601– 638.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock– Johnson Psychoeducational Battery—Third Edition*. Chicago, IL: Riverside.