

Response to Intervention: An Alternative Means of Identifying Students as Emotionally Disturbed

Frank M. Gresham
Louisiana State University
Project REACH
University of California-Riverside

Abstract

Children and youth exhibiting serious emotional, behavioral, and interpersonal problems create substantial challenges for schools, teachers, their parents, and other students. Students having these characteristics are often underserved or unserved by educational and mental health systems in the United States. Recent prevalence rates for children served as emotionally disturbed (ED) under the Individuals With Disabilities Education Act is less than 1 percent although over 20 percent of the school population could qualify for a psychiatric diagnosis. A major reason for the underservice of children as ED lies in the federal definition of emotional disturbance which is nebulous, often illogical, and self-contradictory. An alternative approach to ED identification based on a student's response to an evidence-based intervention is proposed in this article. Response to intervention is defined and described along with methods and procedures for quantifying whether or not a student shows an adequate or inadequate response to an evidence-based intervention implemented with integrity.

• • •

Serious emotional, behavioral, and social difficulties exhibited by children and youth result in substantial challenges to schools, teachers, parents, and their peers. These challenges cut across disciplinary, instructional, and interpersonal domains and often can create chaotic school and classroom environments. A particularly disturbing finding is that students exhibiting severe emotional and behavioral challenges are either underserved or unserved by educational and mental health systems in the United States (National Association of School Psychologists, 2000). Historically, the U.S. Department of Education estimated the prevalence rate for children and youth served as (ED) at 2 percent (Kauffman, 2001). However, recent prevalence estimates of children served as ED continues to be less than 1 percent nationwide (U.S. Department of Education, 2003). Among the states,

Correspondence should be directed to Frank M. Gresham, Graduate School of Education, University of California-Riverside, Riverside, CA 92521
frank.gresham@ucr.edu

the category of ED shows the greatest variability in prevalence of any disability category (Hallahan, Keller, & Ball, 1986). This large degree of variability among states is most likely due to confusion, ambiguities, and differences in the definition and interpretation of ED.

Underservice of Students with ED

The behavioral characteristics and needs of children at-risk for ED has overwhelmed the capacity of schools to effectively accommodate these students. Walker and Gresham (2003) suggested that it is ironic that schools have been slow to recognize the educational needs and demands that these students pose to themselves, to the major social agents in their lives (e.g., parents and teachers), and society at large. Estimates indicate that almost 20 percent of the school-age population could qualify for a psychiatric diagnosis using criteria from the *Diagnostic and Statistical Manual of Mental Disorders-4th Edition* (American Psychiatric Association, 1994; Angold, 2000). Hoagwood and Erwin (1997) suggested that 22 percent of school-age children have mental health problems so severe as to require attention, treatment, and supports. There is a huge disparity between the percentage of children and youth needing mental health services (20+ percent) and those actually served in special education under the Individuals With Disabilities Education Act (IDEA).

Reasons for this underservice appear to be primarily philosophical and fiscal in nature (Walker, Ramsay, & Gresham, 2004). Philosophically, schools have a long history of believing that they are not responsible or accountable for the mental health needs of students. Additionally, the definition of ED in federal legislation (IDEA) has specifically excluded students who are characterized as "*socially maladjusted*." This philosophy is based on the premise that students who have problems in conduct (i.e., social maladjustment) are responsible for their behavior and thus do not have a legitimate disability. In contrast, students who exhibit internalizing behaviors (e.g., anxiety, depression, and fearfulness) do so because these problems are beyond their control. These students are considered to be victims of circumstance and therefore have a "legitimate" disability.

The Issue: What is a "True Disability"?

A major challenge in the identification of students as ED involves a decision regarding whether emotional and/or behavioral difficulties constitute a disability. That is, when does a behavior problem become an "emotional disturbance?" When does social withdrawal and shyness become an anxiety disorder? When does sadness and loneliness become a major depressive disorder? When do overactivity, impulsivity, and inattention become an attention-deficit/hyperactivity disorder? The answer to these questions is not straightforward and ultimately involves some degree of subjective judgment. The category of ED describes a group of students

whose behavior differs from their peers more in terms of *degree* rather than in *kind*. Few individuals would question that children who are profoundly deaf differ in kind from their normally hearing peers in terms of hearing acuity, verbal communication skills, and receptive and expressive verbal language development. There is also little question as to what the definition of profoundly deaf is with respect to the degree of hearing loss (>100 decibels, bilaterally). No such objective tests or criteria exist for determining which students are and are not ED.

The IDEA (1997, 2004) definitions of ED state that it is a condition characterized by one or more of the following characteristics over a long period of time and to a marked degree which adversely affects educational performance: (a) an inability to learn that cannot be explained by intellectual, sensory, or health factors; (b) an inability to build or maintain satisfactory interpersonal relationships with peers or teachers; (c) inappropriate types of behaviors or feelings under normal circumstances; (d) a general pervasive mood of unhappiness or depression; or (e) a tendency to develop physical symptoms or fears associated with personal or school problems. The definition also includes children who are schizophrenic. The definition *excludes* children who are socially maladjusted, unless they are also ED.

A student must meet one or more of the above five criteria to qualify as ED and must also meet all three limiting criteria of severity, duration, and impact on school performance (see Forness & Knitzer, 1992). These limiting criteria, however, are nebulous and highly subjective. Severity derives from the language "to a marked degree." Duration comes from the language of "over a long period of time." Impact is based on the language of "adversely impacts school performance." The most controversial aspect of the ED definition is the *social maladjustment exclusion* clause (see Skiba & Grizzle, 1991). This is discussed below.

The social maladjustment exclusion clause does not allow for students to be deemed eligible as ED if they are socially maladjusted. They can be socially maladjusted, however if they are also ED and therefore receive services. This logic is convoluted, circular, and borders on oxymoronic. The social maladjustment clause in the ED definition excludes and includes a portion of students in the same sentence and directly contradicts several of the five eligibility criteria (Gresham, 1999). An example is the criterion stating that ED is characterized by an inability to build or maintain satisfactory interpersonal relationships with peers or teachers. This criterion essentially defines the concept of social maladjustment (Forness & Knitzer, 1992; Walker et al., 2004). The criterion of inappropriate behavior or feelings under normal circumstances might also be used to characterize the behaviors of many children who are socially maladjusted. In short, the social maladjustment exclusion clause makes no sense in the past and current definitions of ED and is self-contradictory.

Another criticism of the current ED definition involves the impact crite-

tion of "adversely affects educational performance." This language is redundant with "an inability to learn" described earlier in the definition. In addition, *educational performance* has been narrowly construed by most as referring only to academic performance rather than a broader view that includes social, affective, and vocational domains of performance (see National Association of School Psychologists, 2002). There are many children who should qualify for ED, but who do demonstrate adequate and even superior academic performance. A striking example of this is John Nash who was the subject of the book and movie "*A Beautiful Mind*." Nash, a gifted mathematician, won the Nobel Prize for economics and suffered from paranoid schizophrenia his entire life. To say that Nash could not have qualified for ED services because he obviously had superior academic performance would have been ludicrous.

Response to Intervention in EBD Identification and Placement

A relatively new approach to making eligibility determinations as well as selecting or titrating interventions is based on the concept of *response to intervention* (RTI). RTI is based on the logic that if a student's behavioral excesses and/or deficits continue at unacceptable levels subsequent to an evidence-based intervention implemented with integrity, then the student can and should be eligible for ED services (Gresham, 1991, 1999). RTI is based on the best practices of prereferral intervention and gives school personnel the latitude to function within an intervention framework rather than a psychometric eligibility framework.

Definition and Characteristics of RTI

RTI is defined as an inadequate change in target behaviors as a function of intervention. The goal of all interventions is to produce a discrepancy between baseline and post-intervention levels of performance. In fact, within a problem-solving model, a "problem" is defined as a discrepancy between current and desired levels of performance (Bergan & Kratochwill, 1990; Tilly, 2002). The failure to produce a sufficient discrepancy can be taken as partial evidence for an ED eligibility determination. RTI uses data based decision making as a basis for modifying, titrating, or changing the nature of interventions. This logic is not unlike a physician changing the dosage level or type of drug based on the patient's unacceptable response to that drug.

Figure 1 presents a schematic for interpreting the outcomes that might be produced in a RTI approach. The upper left hand quadrant reflects a false positive decision in which a student was identified as ED, but who responded adequately to an intervention. The lower right hand quadrant depicts a false negative decision in which a student was not identified as ED, but who responded adequately to the intervention. The upper right hand quadrant reflects a true positive decision in which a non-responder

to the intervention was identified as ED. The lower left hand quadrant shows a true negative decision in which a responder to the intervention was not identified as ED.

The central issue in using a RTI approach is the determination of "adequate" and "inadequate" response to intervention. This decision must be made at the local and individual level by an assessment and placement team and will most certainly vary across cases and schools that it does now with the current model. This article has detailed a number of ways in which "response to intervention" can be operationalized. The major advantage of adopting a RTI model is that it moves professionals away from addressing the problem to doing something about the problem.

There are a host of factors that are related to a behavior's response to intervention. Several factors that appear to be the most relevant for school-based interventions are: (a) severity of behavior, (b) chronicity of behavior, (c) generalizability of behavior change, (d) treatment strength, (e) treatment integrity, and (f) treatment effectiveness. Each of these factors is discussed in the following sections.

<i>Status</i>	<i>Responder</i>	<i>Non-Responder</i>
	(Not ED)	(ED)
ED	False Positive (Adequate response)	True Positive (Inadequate response)
Not ED	True Negative (Adequate response)	False Negative (Inadequate response)

Figure 1. *Classification of RTI Outcomes.*

Severity. Behavioral severity can be defined using objective dimensions of behavior such as frequency/rate, duration, intensity, and permanent products (Johnston & Pennypacker, 1993). Behavioral severity that is operationalized by high frequencies, durations, and/or intensities is more resistant to intervention than behaviors having lower levels of these behavioral dimensions (Gresham, 1991; Nevin, 1988). These behaviors are not only *more resistant* to interventions but also tend to produce high rates of positive reinforcement (e.g., social attention or access to tangibles) and/or negative reinforcement (e.g., escape or avoidance of task demands) for the student. The net result is that these behaviors continue and even escalate despite interventions designed to reduce them. Using an analogy to physics, the "force" (strength of the intervention) is insufficient to change the "momentum" (severity) of the behavior. Behavioral severity defined in this way meets the IDEA limiting criterion of "to a marked degree."

Chronicity. The chronicity of behavior is an important aspect of almost all classification systems for ED. IDEA requires that behavioral characteristics must have existed "over a long period of time." Many diagnoses in DSM-IV specify that disturbances must have been present for at least six months (e.g., Conduct Disorder, Generalized Anxiety Disorder, Major Depressive Disorder). Thus, the term chronicity implies a condition that is constant, continuing, and of long duration.

Another definition of chronic is "habits that resist all efforts to eradicate them" or "deep-seated aversion to change" (*Webster's New World Dictionary, 2nd College Edition, 1974*). This use of the term chronic is directly related to the concept of response to intervention, or, more accurately, resistance to intervention. Whereas IDEA and DSM-IV use the term chronic as representing constant, continuing, and long duration, a RTI model advocates the second use of the term in defining ED. That is, one distinguishing feature of ED is that it is a pattern of behavior that continues in spite of interventions specifically designed to change it (Gresham, 1991, 1999). Additionally, another use of the term chronic is "the recurrence of behavior problems" once they have been changed by an intervention. This use of the term chronic represents a problem in the *maintenance* of behavior change over time.

Generalizability of behavior change. Generalization and maintenance of behavior change is directly related to RTI. If a behavior pattern is severe (i.e., in terms of frequency, intensity, and/or duration), chronic (i.e., it has been resistant to intervention), it will tend to show *less* generalization across different, non-intervention conditions and will show *less* maintenance over time when intervention procedures are withdrawn (Horner & Billingsley, 1988; Nevin, 1988). Students who demonstrate severe behavior patterns over an extended period of time are quick to discriminate intervention from non-intervention conditions, particularly when intervention conditions are vastly different from non-intervention conditions. For instance, when students are exposed to a highly structured point system that uses a

response cost component for inappropriate behaviors and a reinforcement component for appropriate behaviors, they will readily discriminate when the program is not in effect. Since *discrimination* is the polar opposite of generalization, behavior under these conditions are likely to deteriorate to baseline levels of performance when one returns abruptly to pre-intervention conditions (withdrawal of the point system).

Students with ED often show excellent initial behavior change, particularly in terms of behavioral excesses, but fail to show generalization and maintenance of behavior changes (Horner & Billingsley, 1988). A reason for this lack of generalization and maintenance is that interventions often exclusively target decreasing inappropriate behavioral excesses at the expense of targeting the establishment of appropriate or prosocial behaviors. Furthermore, to ensure generalization and maintenance of intervention effects these effects should be actively programmed (Stokes & Osnes, 1989). Recent advances in *positive behavioral support* in which entire schools recognize and abide by a common set of behavioral expectations for students, for example, should enhance the generalization and maintenance of individualized intervention effects for students with ED (see Sugai, Horner, & Gresham, 2002).

Treatment strength. The strength of a treatment reflects the ability of a given treatment to change behavior in the desired direction. Strong treatments produce greater amounts of behavior change than weak treatments. Treatment strength is not absolute, but rather situationally, behaviorally, and individually specific (Gresham, 1991). Some treatments are strong in some situations or settings, but not others (e.g., home versus school). Some treatments are strong for changing some behaviors, but not others (e.g., work completion versus physical aggression). Some treatments are strong for some individuals, but not other individuals (e.g., students with ED versus students who are not ED). In short, treatment strength is determined by the interaction of situational, behavioral, and individual factors.

In behavioral interventions, treatment strength is not always clearly quantifiable *a priori* as it is in other fields. For example, a 500 mg antibiotic is twice as strong as a 250 mg antibiotic in treating bacterial infections. In contrast, four points awarded in a point system for appropriate behavior is not necessarily twice as strong as two points. The fundamental difference between a specification of treatment strength in medical and behavioral treatments is that the former specifies treatment strength *a priori* (e.g., dosage of drug) and the latter specified treatment strength *a posteriori* (e.g., magnitude of behavior change). Treatment strength in a RTI model is indexed by treatment outcome or magnitude of behavior change produced by a treatment. The criteria used to judge whether or not a treatment is effective is discussed later in this article.

Treatment integrity. The degree to which a given treatment is implemented as planned or empirically validated describes the concept of treatment integrity (Gresham, 1989, 1997). Treatment integrity involves the ac-

curacy and consistency with which an intervention is implemented. Treatment integrity is an essential ingredient in a RTI model because effective treatments can be rendered ineffective simply because they were either implemented poorly or not implemented at all. It is also possible for ineffective treatments to be implemented with perfect integrity but have no effect on behavior change. Many academic and behavioral interventions designed in a consultative relationship produce ineffective results because of the poor integrity with which these interventions are delivered (Noell & Witt, 1999).

The level of integrity with which behavioral interventions are implemented in applied settings is likely to be lower than what is reported in the research literature. The integrity of interventions depends on several factors such as the complexity of the intervention (Gresham, 1989; Yeaton & Sechrest, 1981), the time required to implement the intervention (Noell & Gresham, 1993), the materials and resources required to implement the intervention correctly (Gresham, 1989; Woodward & Gersten, 1992), and the perceived and actual effectiveness of the intervention (Elliott, 1988; Elliott, Witt, Kratochwill, & Stoiber, 2002). Progress in using a RTI model will require systems for measuring and enhancing the integrity of interventions delivered in school settings.

Treatment effectiveness. The conceptualization of ED presented in this article requires that a school-based intervention be implemented with integrity and fail to show an adequate effect on behavior *prior* to an eligibility determination decision. In short, if a behavior pattern continues at an unacceptable level (i.e., the behavior pattern does not respond to the intervention), then an eligibility determination of the student as ED might be warranted. In a RTI model, how does one know whether or not a given treatment was effective in changing a pattern of behavior? What standards or criteria might one use to make this decision? Four approaches that have been proposed to quantify whether or not treatments are effective are described: (a) visual inspection of data, (b) reliable changes in behavior, (c) changes on social impact measures, and (d) social validation.

Visual inspection. To determine whether or not an individual's behavior has changed as a function of intervention, one can use *visual inspection* of graphed data for a single individual from baseline to intervention phases. Intervention effects are determined by comparing baseline levels of performance to post-intervention levels of performance. Unlike traditional statistical analyses, visual inspection relies on the "interocular" test of significance. The logic of visual inspection is quite simple: If a meaningful effect was produced by the treatment, it should be obvious or noticeable by simply viewing graphed data (Baer, 1977; Morgan & Morgan, 2001). Potential drawbacks of relying exclusively on visual inspection include: (a) an absence of standards or benchmarks for deciding if behavior change is clinically or educationally significant, (b) potential for unacceptably high Type I error rates, and (c) difficulty in interpreting autocorrelated time series data.

Reliable changes in behavior. An essential requirement of a RTI model is that it must be demonstrated that changes in behavior produced by an intervention are *reliable changes* and are not due to chance or extraneous factors. Five metrics have been proposed to quantify the extent to which changes in behavior are reliable: (a) absolute change indices (Kazdin, 2003), (b) reliable change indices (Christensen & Mendoza, 1986; Jacobson, Follette, & Revenstorf, 1984), (c) percent nonoverlapping data points (Mastropieri & Scruggs, 1985-86), (d) percent change from baseline, and (e) effect size estimates (Busk & Serlin, 1992). Each of these is described briefly in the following paragraphs.

Absolute change is the degree or amount of change an individual makes that does not involve comparison to other groups (Kazdin, 2003). Absolute change can be calculated in one of three ways: (a) the amount of change from baseline to post-intervention levels of performance, (b) an individual no longer meeting established criteria for ED, and (c) the total elimination of behavior problems. Absolute change is straightforward, intuitively logical, an easy to calculate. It is also consistent with a problem-solving approach to defining behavior problems as the discrepancy between expected and desired levels of performance described earlier. Using this approach, a problem is considered "solved" if the degree of absolute change is large relative to the three criteria described above.

There are some problems with using metrics of absolute change. For example, an individual might show a relatively large amount of change from baseline to post-intervention levels of performance, but this change might not be large enough to allow that individual to function within a general education setting. Absolute change interacts with tolerance levels for problem behavior at the classroom and school levels. That is, even though a change in behavior is large, the behavior pattern still might not be tolerated by significant others in the school environment. Also, an individual may no longer meet the diagnostic criteria for ED, but this may be due to biases operating in the diagnostic and eligibility decision-making process.

The *reliable change index* (RCI) is calculated by subtracting an individual's posttest score on an outcome measure from his/her pretest score and dividing this difference by the standard error of difference between post- and pre-test scores (Nunnally & Kotsche, 1983). The standard error of difference represents the variability in the distribution of change scores that would be expected if no change had occurred. A RCI of $+1.96$ ($p < .05$) would be considered a reliable change in behavior.

The RCI metric has the advantage of quantifying reliable changes from baseline to post-intervention levels of performance and confidence intervals can be placed around change scores to avoid overinterpretation of results. The RCI, however, is affected by the reliability of the outcome measures used. For example, if a measure is highly reliable (.90 or greater), then small changes in behavior might be considered statistically reliable, but not socially or clinically important. In contrast, if a measure has rela-

tively low reliability, then large changes in behavior may be socially important, but not considered statistically reliable (Gresham & Noell, 1993).

In addition, the interpretation of RCI is clouded when using direct observational measures (e.g., frequency or duration measures) because "reliability" has a different meaning for these measures. In direct observational data, "reliability" is typically calculated by interobserver agreement indices (e.g., percent agreement on occurrence/nonoccurrence of behavior). This is not the same as reliability in the traditional use of the term (see Gresham, 2003 for a discussion). As such, calculation of RCI does not have the same meaning because the data are nonparametric (they are not based on a normal distribution of test scores).

Percent nonoverlapping data points (PND) is a metric computed by calculating the percentage of nonoverlapping data points between baseline and intervention phases (Mastropieri & Scruggs, 1985-86). If the goal is to decrease problem behavior, one computes PND by counting the number of intervention data points exceeding the *highest* baseline data point and dividing by the total number of data points in the intervention phase. For example, if 9 of 10 treatment data points exceed the highest baseline data point, the PND would be 90%. Alternatively, if the goal is to increase behavior (e.g., social skills), then one calculates PND by counting the number of intervention data points that are below the *lowest* baseline data point and dividing by the total number of data points in the intervention phase. PND provides a quantitative index to document the effects on an intervention that is easy to calculate. There are, however, some drawbacks of using this method that should be noted. One, PND often does not reflect the magnitude of effect an intervention. That is, one can have 100% nonoverlapping data points from in the treatment phase yet have an extremely weak treatment effect. Two, unusual baseline trends (high and low data points) can skew the interpretation of PND. Three, PND is greatly affected by floor and ceiling effects. Four, aberrant or outlier data points can make interpretation of PND difficult (see Strain, Kohler, & Gresham, 1998 for a discussion). Five, there are no well-established empirical guidelines for what constitutes a large, medium, or small effect using the PND metric.

An alternative to the PND statistic is to calculate the *percent change* in behavior from baseline to post-intervention levels of performance. This metric involves comparing the mean level of performance in baseline to the mean level of performance in intervention. For example, if the mean frequency of a behavior in baseline were 8 and the mean frequency of behavior after intervention was 2, then the percent change in behavior would be 75% ($(8-2)/8 = .75$ or 75%). The advantage of the percent change metric is that outliers or aberrant data points or floor and ceiling effects do not as greatly affect it as the PND index. Percent change in behavior is commonly used in medicine to evaluate the effects of medical treatments such as drugs that reduce cholesterol or blood pressure. There are well-established medical benchmarks for desirable levels of blood cholesterol

(<200 *dl*) and blood pressure (120/80). Unfortunately, there are no such adequate benchmarks for many behaviors targeted for intervention in the ED population. Also, like the PND, there are no clear guidelines for determining the magnitude of behavior change that is sufficient to indicate an individual has demonstrated an adequate response to intervention. As such, this metric must be supplemented by other measures (e.g., social validation measures) that are described in subsequent sections of this article.

Effect size. Effect size estimates for the individual case is a modification of the effect size estimate known as Cohen's *d* that is used with group meta-analytic research. Busk and Serlin (1992) proposed two methods for calculating effect sizes at the individual level. The first method makes no assumptions about the distribution of data points in baseline and intervention phases. It is calculated by subtracting the intervention mean from the baseline mean and dividing by the standard deviation of the baseline mean. The second approach makes an assumption with respect to homogeneity of variance in the data points and uses the pooled standard deviation calculated from baseline and intervention phases in the denominator. A drawback of using this latter effect size estimate is that it can yield large effect size estimates than cannot be interpreted in the same way as effect sizes calculated from group intervention data.

Changes on social impact measures. The ultimate goal in intervention for students who are at-risk for ED is to change their standing on measures of social impact. A social impact measure is characterized by changes that are recognized by most as being critically important in everyday life (Kazdin, 2003). These measures represent socially valued intervention goals because social systems such as schools and mental health agencies utilize them to index the success or failure of interventions (Gresham, 1983). Examples of social impact measures include: school dropout, arrest rates, days missed from school, and school suspensions/expulsions. These measures might be considered criterion measures against which behavior changes can be validated.

The drawback in using social impact measures is that they are not particularly sensitive in detecting short-term intervention effects. Many treatment consumers consider these social impact measures to be the bottom line in gauging successful intervention outcomes, however exclusive reliance on these measures might ignore a great deal of behavior change (Kazdin, 1992). As such, exclusive reliance on social impact measures in a RTI model may result in unacceptably high Type II error rates (retaining a false null hypothesis).

It is often the case that rather large and sustained changes in behavior are required before these changes are reflected on social impact measures. Sechrest, McKnight, and McKnight (1996) suggested using the method of *just noticeable differences* (JND) to index intervention outcomes. The JND approach answers the question: How much of a difference in behavior is required before it is "noticed" by significant others or reflected on social

impact measures? In the case of a student at-risk for ED, how much of a decrease in aggressive/disruptive behavior is required before it is reflected in a decrease and subsequent elimination of office discipline referrals?

Social validation. Social validity addresses three fundamental questions asked by professionals concerned with ED: What should we change? How should we change it? How will we know it was effective? There are often disagreements among professionals and between professionals and treatment consumers on these three fundamental questions (Hawkins, 1991; Schwartz & Baer, 1991). Wolf (1978) described the social validation process as involving the assessment of the *social significance* of intervention goals, the *social acceptability* of intervention procedures, and the *social importance* of intervention effects. This last aspect of the social validation process is the most relevant in quantifying treatment effectiveness in a RTI model.

Establishing the social importance of the effects of an intervention attests to the practical or educational significance of behavior change for the student. Do the quantity and quality of the changes in behavior make a difference in the student's behavioral functioning and adjustment? In short, do the changes in behavior have *habilitative validity* (Hawkins, 1991)? Is the student's behavior now in the functional range subsequent to the intervention? These questions capture the essence of establishing the social importance of intervention effects.

A way of establishing the social importance of intervention effects is to view behavioral functioning as belonging to either a functional or dysfunctional distribution. An example might be socially validating a behavioral intervention by showing the student's behavior moved from a dysfunctional to a functional range of performance. Using teacher and parent ratings on nationally normed behavior rating scales is a means of quantifying social importance of intervention effects (Gresham & Lopez, 1996). Moving a student's problem behavior ratings from the 95th percentile to the 50th percentile would represent a socially important change. Similarly, changing the target behavior problem measured by direct observations into the range of non-referred peers would also corroborate the behavior ratings and therefore could be considered socially important.

The social importance of effects is perhaps best conceptualized and evaluated on several levels: proximal effects, intermediate effects, and distal effects (Fawcett, 1991). Proximal effects are changes in target behaviors produced by the intervention such as increases in social skills, decreases in aggressive behavior, or decreased anxiety. Proximal effects can be evaluated using visual inspection of graphed data or percent change in behavior from baseline to intervention. Intermediate effects can be evaluated by more molar assessments such as substantial changes in ratings on normed behavior rating scales. Distal effects can be evaluated by changes on social impact measures such as office discipline referrals, suspension/expulsion rates, school attendance, reincarceration rates, or parole violators.

Intervention Principles

An essential requirement in using a RTI approach is that there must be validated intervention protocols and procedures to change behavior. There are a number of validated interventions that can be implemented to change the behavior of students at-risk for ED. These have been comprehensively described under the rubric of *evidence-based* interventions in both school psychology (Kratochwill & Stoiber, 2000; Stoiber & Kratochwill, 2000) and clinical child psychology (Gresham, Cook, Crews, & Kern, 2004; Lonigan, Elbert, & Bennett-Johnson, 1998; Weisz, Weiss, Alicke, & Klotz, 1987) literatures. School-based interventions are often conceptualized on three levels: *universal interventions*, *selected interventions*, and *targeted/intensive interventions* (Sugai et al., 2002). An important concept in a RTI model is that it matches the intensity of the intervention to both the severity of the problem behavior and the problem behavior's resistance to intervention efforts.

Universal interventions are delivered to all students under the same conditions and are implemented at a district wide, school wide, or classroom wide levels. It is estimated that approximately 80-90 percent of any given school population will respond adequately to universal interventions (Gresham, 2004; Sugai et al., 2002; Walker et al., 2004). Selected interventions are a class of interventions that focus on the nonresponders to universal interventions. These students typically are at greater risk for severe problem behaviors and will require more intensive intervention resources. Inadequate responders may respond to relatively simple individually focused interventions such as social skills interventions, token systems, behavioral contracts, or self-management strategies (Gresham, 2004; Gresham et al., 2004; Walker et al., 2004).

Targeted/intensive interventions represent the most intense level of intervention and target students with the most severe and resistant behaviors. Many students served under the category of ED will require this level of intervention. Estimates suggest that these students constitute about 1-5 percent of a given school population, they account for 40-50 percent of behavioral disruptions in schools, and they drain 50-60 percent of school building and classroom resources (Colvin, Kame'enui, & Sugai, 1993; Gresham, 2004; Sugai et al., 2002). These students will require the most intense, individualized, and comprehensive system of intervention supports involving multiple social agencies such as mental health, juvenile justice, and social services (Walker et al., 2004).

Conclusion

Many children and youth who might otherwise qualify for special education and related services under the category of ED are not identified as such and therefore do not receive the most appropriate education. The

degree of underservice for the potential ED population is huge with less than 1 percent of students nationwide being served as ED. Despite this, over 20 percent of the school age population has emotional and behavioral difficulties severe enough to qualify for psychiatric diagnosis. A major reason for this underservice lies in the various ways in which the concept of ED has been defined.

The definition of ED used by schools is based on the federal definition found in IDEA. This definition was criticized as being nebulous and self-contradictory. A particularly troublesome aspect of the definition is and always has been the *social maladjustment exclusion* clause. This clause is virtually uninterpretable because it excludes children who are socially maladjusted unless they are also ED. Also, the definition was criticized because it requires that the emotional disturbance adversely affect educational performance. This has been narrowly interpreted by most local education agencies to mean low academic achievement. "Educational performance" has recently been broadened to include social, affective, and vocational domains (National Association of School Psychologists, 2002).

An alternative definition that is consistent with a problem solving and data-based decision-making approach was described in this paper as the RTI model. In a RTI model, it was suggested that a student should be considered for ED services if the student's behavior does not change adequately to an evidence-based intervention implemented with integrity (Gresham, 1991, 2004). Unlike current practices, this approach to the identification of ED requires implementation of an intervention prior to making an eligibility determination. Current practice now is based on a refer-test-place model in which students are not exposed to systematic, evidence-based interventions to ameliorate behavior problems. The RTI model is seen as an improvement over the current practice of eligibility determination that excludes or defers best practice interventions.

References

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Angold, A. (2000). *Preadolescent screening and data analysis*. Paper presented to the 2nd Annual Expert Panel Meeting on Preadolescent Screening Procedures. Sponsored by the Substance Abuse and Mental Health Services Administration, National Institutes of Health, Washington, DC.
- Baer, D.M. (1977). "Perhaps it would be better not to know everything." *Journal of Applied Behavior Analysis*, 10, 167-172.
- Bergen, J., & Kratochwill, T.R. (1990). *Behavioral consultation and therapy*. New York: Plenum Press.
- Busk, P., & Serlin, R. (1992). Meta-analysis for single-case research. In T. Kratochwill & J. Levin (Eds.), *Single-case design and analysis* (pp. 187-212). Hillsdale, NJ: Lawrence Erlbaum.
- Christensen, L., & Mendoza, J. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, 17, 305-308.
- Colvin, G., Kame'enui, E.J., & Sugai, G. (1993). Schoolwide and classroom management: Reconceptualizing the integration and management of students with behavior problems in general education. *Education and Treatment of Children*, 16, 361-381.

- Elliott, S.N. (1988). Acceptability of behavioral treatments in educational settings. In J.C. Witt, S.N. Elliott, & F.M. Gresham (Eds.), *Handbook of behavior therapy in education* (pp. 121-150). New York: Plenum.
- Elliott, S.N., Witt, J.C., Kratochwill, T.R., & Stoiber, K.C. (2002). Selecting and evaluating classroom interventions. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 243-294). Bethesda, MD: National Association of School Psychologists.
- Fawcett, S. (1991). Social validity: A note on methodology. *Journal of Applied Behavior Analysis*, 24, 235-239.
- Forness, S., & Knitzer, J. (1992). A new proposed definition and terminology to replace "serious emotional disturbance" in Individuals with Disabilities Education Act. *School Psychology Review*, 21, 12-20.
- Gresham, F.M. (1983). Social validity in the assessment of children's social skills: Establishing standards for social competency. *Journal of Psychoeducational Assessment*, 1, 299-307.
- Gresham, F.M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37-50.
- Gresham, F.M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. *School Psychology Review*, 20, 23-36.
- Gresham, F.M. (1997). Treatment integrity in single-subject research. In R. Franklin, D. Allison, & B. Gorman (Eds.), *Design and analysis of single-case research* (pp. 93-118). Mahwah, NJ: Lawrence Erlbaum.
- Gresham, F.M. (1999). Noncategorical approaches to K-12 emotional and behavioral difficulties. In D. Reschly, W.D. Tilly, & J. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 107-138). Longmont, CO: Sopris West.
- Gresham, F.M. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review*, 33, 326-343.
- Gresham, F.M. (2003). Establishing the technical adequacy of functional behavioral assessment: Conceptual and measurement challenges. *Behavioral Disorders*, 28, 282-298.
- Gresham, F.M., & Lopez, M.F. (1996). Social validation: A unifying concept for school-based consultation research and practice. *School Psychology Quarterly*, 11, 204-227.
- Gresham, F.M., & Noell, G.H. (1993). Documenting the effectiveness of consultation outcomes. In J. Zins, T. Kratochwill, & S. Elliott (Eds.), *Handbook of consultation services for children* (pp. 249-276). San Francisco: Jossey-Bass Publishers.
- Gresham, F.M., Cook, C.R., Crews, S.D., & Kern, L. (2004). Social skills training for children and youth with emotional and behavioral disorders: Validity considerations and future directions. *Behavioral Disorders*, 30, 19-33.
- Hallahan, D.P., Keller, C.E., & Ball, D.W. (1986). A comparison of prevalence rate variability from state to state for each of the categories of special education. *Remedial and Special Education*, 7, 8-14.
- Hawkins, R.P. (1991). Is social validity what we are interested in? Argument for a functional approach. *Journal of Applied Behavior Analysis*, 24, 205-213.
- Hoagwood, K., & Erwin, H. (1997). Effectiveness of school-based mental health services for children: A 10-year research review. *Journal of Child and Family Studies*, 6, 435-451.
- Horner, R., & Billingsley, F. (1988). The effects of competing behavior on the generalization and maintenance of adaptive behavior in applied settings. In R. Horner, G. Dunlap, & R. Koegel (Eds.), *Generalization and maintenance: Lifestyle changes in applied settings* (pp. 197-220). Baltimore: Brookes.
- Individuals with Disabilities Education Act, 20, U.S.C. Ch. 33, Sec 1400, (1997).
- Jacobson, N., Follette, W., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336-352.
- Johnston, J., & Pennypacker, H. (1993). *Strategies and tactics of human behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Kauffman, J.M. (2001). *Characteristics of emotional and behavioral disorders of children and youth* (7th ed.). Upper Saddle, NJ: Merrill Prentice-Hall.

- Kazdin, A.E. (1992). *Research design in clinical psychology* (2nd ed.). New York: Macmillan Publishing Co.
- Kazdin, A.E. (2003). Clinical significance: Measuring whether interventions make a difference. In A.E. Kazdin (Ed.), *Methodological issues & strategies in clinical research* (3rd ed., pp. 691-710). Washington, DC: American Psychological Association.
- Kern, L., & Gresham, F.M. (2002). *Center for Students Requiring Intensive Social, Emotional, and Behavioral Interventions: Project REACH*. U.S. Department of Education, Office of Special Education Programs.
- Kratochwill, T.R., & Stoiber, K.C. (2000). Empirically supported interventions and school psychology: Conceptual and practice issues—Part II. *School Psychology Quarterly*, 15, 233-253.
- Lonigan, C.J., Elbert, J.C., & Bennett Johnson, S. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology*, 27, 138-145.
- Mastropieri, M., & Scruggs, T. (1985-86). Early intervention for socially withdrawn children: *Journal of Special Education*, 19, 429-441.
- Morgan, D.L., & Morgan, R.K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist*, 56, 119-127.
- National Association of School Psychologists (2002). *Position statement on students with emotional and behavioral disorders*. Bethesda, MD: Author.
- Nevin, J. (1988). Behavioral momentum and the partial reinforcement effect. *Psychological Bulletin*, 103, 44-56.
- Nunnally, J., & Kotsche, W. (1983). Studies of individual subjects: Logic and methods of analysis. *British Journal of Clinical Psychology*, 22, 83-93.
- Schwartz, I., & Baer, D.M. (1991). Social validity assessment: Is current practice state of the art? *Journal of Applied Behavior Analysis*, 24, 189-204.
- Sechrest, L., McKnight, P., & McKnight, K. (1996). Calibration of measures for psychotherapy outcome studies. *American Psychologist*, 51, 1065-1071.
- Skiba, R., & Grizzle, K. (1991). The social maladjustment exclusion: Issues in definition and assessment. *School Psychology Review*, 20, 23-28.
- Stoiber, K.C., & Kratochwill, T.R. (2000). Empirically supported interventions and school psychology: Rationale and methodological issues—Part I. *School Psychology Quarterly*, 15, 75-105.
- Stokes, T., & Osnes, P. (1989). An operant pursuit of generalization. *Behavior Therapy*, 25, 429-432.
- Strain, P., Kohler, F., & Gresham, F.M. (1998). Problems in logic and interpretation of quantitative synthesis of single case research: Mather and colleagues as a case in point. *Behavioral Disorders*, 24, 74-85.
- Sugai, G., Horner, R.H., & Gresham, F.M. (2002). Behaviorally effective school environments. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (2nd ed., pp. 315-350). Bethesda, MD: National Association of School Psychologists.
- Tilly, W.D. (2002). Best practices in school psychology as a problem solving enterprise. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 21-36). Bethesda, MD: National Association of School Psychologists.
- U.S. Department of Education (2003). *Twenty-fifth annual report to Congress on implementation of the Individuals With Disabilities Education Act*. Washington, DC: Author.
- Walker, H.M., & Gresham, F.M. (2003). School-related behavior disorders. In W.M. Reynolds & G.E. Miller (Eds.), *Handbook of psychology: Educational psychology* (vol. 7, pp. 511-530). New York: John Wiley & Sons.
- Walker, H.M., Ramsay, E., & Gresham, F.M. (2004). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Weisz, J., & Weiss, B., Alicke, M.D., & Klotz, M.L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology*, 55, 542-549.

- Wolf, M.M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203-214.
- Woodward, J., & Gersten, R. (1992). Innovative technology for secondary learning disabled students: A multi-faceted study of implementation. *Exceptional Children, 58*, 407-421.
- Yeaton, W., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49*, 156-167.

Copyright of Education & Treatment of Children is the property of ETC. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.