The Use of Change-Sensitive Measures to Assess School-Based Therapeutic Interventions:
Linking Theory to Practice at the Tertiary Level

Amanda L. Lannie
Devereux Center for Effective Schools
Robin S. Codding
University of Massachusetts, Boston
James L. McDougal
State University of New York at Oswego
Scott Meier
State University of New York at Buffalo

ABSTRACT: Students with chronic problem behavior will likely require intense support services, often referred to as tertiary-level supports. This article highlights the need for change-sensitive measures to guide the evaluation of tertiary-level supports in the schools (e.g., therapeutic services). In this regard, we propose a model for using change-sensitive measures to document student progress at a programmatic level. Paralleling the decision-making process utilized in a positive behavior interventions and supports (PBIS) approach, we propose a model for tertiary-level service evaluation that relies in part on providers collecting change-sensitive data and generating graphs to aid in data-based decision making practices. Similar to the PBIS process, data to be graphed would span five key dimensions. Last, we present a case illustration where data are analyzed in accordance with the proposed model to assess the services at a programmatic level.

Prevalence estimates of emotional–behavioral disorders among children and youth range, but most agree that about 20% of children present with diagnosable disorders (i.e., U.S. Department of Health and Human Services, 1999) and a conservative prevalence of 3–6% of children presenting with serious and chronic disorders (Kauffman, 1997). Disorders of this magnitude do typically evolve in a predictable progression first manifesting in the form of either behavioral excess or behavioral deficits (externalizing or internalizing problems) during childhood. Up to 75% of children demonstrating significant externalizing behaviors (severe tantrums, disobedience) eventually engage in predictable and serious law breaking and antisocial behavior (e.g., Reid, 1993). Internalizing disorders too are quite stable, with children demonstrating significant internalizing difficulties (anxiety, depression) also likely to encounter increased rates of pathology and lower rates of socialization and academic attainment (Hops, Walker, & Greenwood, 1988).

The Surgeon General's report on mental health (U.S. Department of Health and Human Services, 2000) highlighted the urgent need for more accessible, effective, and outcome-oriented services for

children and youth to reduce or prevent this developmental progression of disorder. School-based and school-linked mental health services have been implemented across the country, with school psychologists often leading the efforts, as approaches to meet the needs of this population with a delivery system that is both accessible to youth and sustainable across time (e.g., Atkins, Graczyk, Frazier, & Abdul-Adil, 2003). While these school-related services have increased access for children, their effectiveness has not yet been adequately demonstrated (Pelham & Massetti, 2003).

POSITIVE BEHAVIOR INTERVENTION AND SUPPORTS

Additionally, several comprehensive, research-based models for providing emotional and behavioral support to students in the schools have evolved in the literature. Perhaps the most widely implemented is school-wide positive behavior support (SWPBS; e.g., Horner, Sugai, Todd, & Lewis-Palmer, 2005). SWPBS, a systems-level application of positive behavior interventions and supports (PBIS), was designed by behavioral and educational professionals to support schools to more effectively promote positive student behavior, decrease antisocial behavior, and support students with emotional and behavioral disorders (e.g., Eber, Sugai, Smith, & Scott, 2002; Netzel & Eber, 2003). SWPBS is a three-tiered data-driven approach that utilizes a team-based collaborative and systemic approach to designing, implementing, and evaluating school-wide behavioral interventions (Horner et al.). Relying on the use of available school data for intervention design, monitoring, and evaluation, a PBIS school team engages in planning across three levels of prevention and intervention aims. The three tiers, modeled after the public health approach, comprise primary (or universal), secondary, and tertiary prevention efforts (Horner et al.). At each tier, interventions increase in intensity to match students' needs with the ultimate goal of supporting all students via efficient and effective means. Prevention practices at the universal level include establishing and teaching school-wide rules, as well as structured systems of recognition of positive behavior and systematic approaches to responding to problem behavior (Horner et al.). Interventions at the secondary level are typically group based in nature to address those students who are not responding to the universal level. At the tertiary level are interventions for those students with severe, chronic patterns of behaviors. These students are best served with an individualized, intensive level of service to adequately address their behavioral needs (Walker et al., 1996). Due to its demonstrated effectiveness, PBIS has recently been offered as a model for school-based mental health (Kutash, Duchnowski, & Lynn, 2006). With functional behavioral assessment (FBA) as the cornerstone of services at the tertiary level, PBIS aims to address challenging behaviors through function-based, contextually relevant, and culturally competent interventions for those students who may indeed present with a diagnosable mental health disorder.

At the universal level, PBIS teams include representatives across specialization areas, grade levels, and administration and meet on a routine basis to examine school-level data to inform intervention development, implementation, and evaluation activities. Specifically, five main graphs are used to focus intervention design with respect to problematic behaviors, students, or settings and additionally to assess preintervention and postintervention change. Continuous monitoring (e.g., monthly) allows for timely intervention revision or intensification, and can be applied to primary and secondary levels of prevention programming.

DATA-BASED DECISION MAKING IN PBIS

The Office of Special Education Programs (2004) Technical Assistance Center on Positive Behavioral Interventions and Supports has identified a number of potential data sources available to schools, including academic scores, attendance, direct behavioral observations, and suspension rates. While a number of possibilities are listed, emphasis in this model is generally focused on the use of office discipline referrals (ODRs) as the primary ongoing barometer of student behavior to be used by school teams when developing and assessing interventions (Sugai, Sprague, Horner, & Walker, 2000; Tobin, Sugai, & Colvin, 2000). ODRs are defined as an instance where a student is referred to the office for a violation of school rules and the severity was beyond that which could be addressed expeditiously by the referring teacher and which

resulted in the completion of an office referral form (e.g., Tobin & Sugai, 1999). ODRs have been used in the SWPBS literature in several ways: (a) to guide intervention selection or design at the universal, targeted, and indicated levels of programming; (b) to monitor or gauge the effectiveness of these intervention programs; and (c) as a procedure for screening students to identify those in need of increasingly intense levels of intervention (see Clonan, McDougal, Clark, & Davison, 2007).

As described, at the universal level the primary source of evaluation data is ODRs, which are used to target areas for intervention and to assess the impact of implemented interventions. It is recommended that PBIS teams convene at least monthly to review ODR data across five core domains: (a) average ODRs per day by month, (b) number of ODRs by problem behavior, (c) number of ODRs by location, (d) number of ODRs by time, and (e) number of ODRs by student (Horner et al., 2005).

While ODRs may be adequate for implementing school-wide programs of behavioral support, they are inadequate for monitoring the emotional and behavioral progress of students displaying clinical level symptoms (i.e., those students typically served at the tertiary level of a SWPBS process; see Chafouleas, Riley-Tillman, & Sugai, 2007; Morrison & Skiba, 2001; Nelson, Martella, & Marchand-Martella, 2002; Wright & Dusek, 1998). The use of ODRs for screening purposes (e.g., Nelson, Benner, Reid, Epstein, & Currin, 2002) and as predictors of school violence (e.g., Morrison & Skiba, 2001) has also been questioned in the literature. The data-based formative assessment/decision-making practices embedded within SWPBS, however, where program implementers review a consistent set of graphs to aid in data-based decision making, would seem to have merit for the evaluation of therapeutic services provided in the schools.

BARRIERS TO ASSESSING TERTIARY-LEVEL INTERVENTIONS

School psychologists are often charged with the task of providing therapeutic services to students in schools. However, the lack of feasible and appropriate outcome-measurement devices (which is also true in adult populations) makes the evaluation of such services a significant barrier (e.g., Meier, 1997). The problem with traditional measures is that, historically, educational and psychological tests were designed to select individuals for entrance to educational or vocational settings or to assign them to a variety of diagnostic groupings. Therefore, standard test-development procedures emphasize the selection of items that discriminate among individuals on traits of interest. These trait-related items are selected to maximize variability among individuals and to demonstrate high stability over time. Further, in order to increase the reliability and validity coefficients of tests composed of trait-based items, developers typically aggregate these measures across items, individuals, and occasions (time). These characteristics result in measures that discriminate among individuals on important traits and yet limit the utility of these tests for assessing intra-individual "state-related" change over time, as well as impeding formative assessment where the focus is on assessing student change over time.

The construct of interest in this area is *therapeutic change* or the specific intra-individual changes that occur for the client (student, consumer) during and after the therapeutic process. Research in the area of counseling and psychotherapy outcomes indicates that the amount of client change assessed can be significantly influenced by the specific method or measure chosen. In fact, researchers have found that there are reliable differences in the sensitivity to change across common outcome measures and that these rating devices can by themselves produce differences larger than those typically attributed to treatments (i.e., Lambert, 1994).

INTERVENTION ITEM SELECTION RULES

Drawing on concepts described by criterion-referenced and longitudinal test developers, Meier has developed a model for test construction based on the selection of intervention-sensitive items. Like traditional items, *intervention-sensitive* items should be theoretically based, be unrelated to systematic error sources, and should avoid ceiling and floor effects. In addition, intervention-sensitive items should demonstrate unique properties

in that they should change in response to an effective intervention and yet remain stable over time when no intervention is present. Thus, scores on these items should show (a) no pretest differences between treatment and similar control groups, (b) pre–post differences in theoretically expected directions for treatment groups, and (c) no pre–post differences for the control group. The nine-step model for selecting intervention-sensitive items and the literature supporting this model are briefly described below.

Meier (1997, 1998) proposed a model for test construction based on a nine-step process called intervention item selection rules (IISRs). This process for identifying intervention-sensitive items is based on two theoretical suppositions: (a) it is useful to consider test items and tasks as differing along a trait–state continuum and (b) different test construction and item analysis procedures are necessary to select items with a high state loading, to more sensitively and accurately reflect the results of psychosocial interventions. Within the IISR model, intervention-sensitive items are selected based on a stringent nine-step process. Selected items should (a) be grounded in theory (to provide for interpretation and minimize the effects of chance), (b) be aggregated across individuals (to reduce the effects of random error), (c) avoid ceiling and floor effects, (d) evidence change after a psychosocial intervention, (e) evidence change in the theoretically expected direction, (f) evidence change in relation to control and comparison groups, (g) show no difference between comparison groups prior to intervention, (h) be unrelated to systematic error sources, and (i) be subject to cross validation studies.

The constructs measured by IISR scales are psychological/psychosocial states (such as depression), or state aspects of a phenomenon or ability (i.e., ability to maintain attention or focus). The construct of intervention- or therapy-related change is operationalized on the scales composed of these items as pre–post changes over time. IISR procedures produce scales with scores that reflect influences, including the effects of situations, environments, and interventions. As previously indicated, scales composed of items with high state loadings have been found to more accurately reflect the results of psychosocial interventions. These items are less likely to be stable over time than the trait-loaded items selected by traditional test-development procedures. The IISR items are, however, more sensitive to intervention effects and may provide more specificity about the effects of an intervention (e.g., less fidgeting and anxiety) as opposed to more global scales (e.g., less dysfunction).

Meier has studied scales constructed with both traditional and IISR procedures in a variety of clinical and school settings (i.e., Meier, 1998, 2000, 2004). Overall, scales constructed with IISR procedures demonstrated larger treatment effect sizes than traditional scales and adequate reliability estimates.

Meier and colleagues created a scale called the Behavior Intervention Monitoring and Assessment System (BIMAS) based on IISR procedures and the results of earlier studies. BIMAS items were selected based on (a) a review of the counseling and psychotherapy outcome literature; (b) examination of item change characteristics in published studies and unpublished program evaluations; and (c) item content and wording suggestions by counselors, psychologists, and social workers in school and clinic settings. BIMAS items were initially developed on the basis of a behavior problem categorization by Stiffman et al. (1984). The review by Stiffman et al. indicated that children's behavior problems could be classified along 13 dimensions, 10 of which were deemed relevant for outcome assessment. These 10 dimensions are sharing thoughts with others, bizarre behaviors, paying attention, impulsive behaviors (not finishing tasks), aggressive behaviors, social skills, anxiety and arousal, physical illness, lying and cheating, and controlling one's temper. Items assessing these dimensions were developed and included in the construction of the initial BIMAS drafts.

Following a series of multiple-year pilot studies, outcome data were collected and analyzed using IISR procedures (Meier, 1998, 2000, 2004; Meier, McDougal, & Bardos, 2008). These analyses, in addition to expert opinion regarding item content, led to the development of the current BIMAS. In addition, items included in the current version are categorized in the following empirically based composites:

- Cognitive processing: Attention, focus, bizarre thoughts/behaviors
- Externalizing behaviors: Conduct problems, substance abuse, deviance

- Internalizing behaviors: Anxiety, depression
- Adaptive functioning: Academic, social, communication

The BIMAS is a 19-item rating scale consisting of 11 items that rate strengths and 8 items that rate problem behaviors. There are three parallel forms designed to collect ratings from parents, teachers, and clinicians. Informants are asked to rate students on a scale from 0 to 3 to indicate whether a student never, rarely, sometimes, or often engages in a behavior. Items are reverse scored to provide ratings for strength items. Therefore, higher scores indicate problems that occur sometimes or often and strengths that occur never or rarely.

Scores on this scale comprise a total score and four subscale scores: (a) adaptive, (b) cognitive/attention, (c) externalizing, and (d) internalizing. Items 2, 4, 5, 11, 12, 14, 17, and 19 are summed to yield the adaptive score. Items 1, 3, 6, and 13 are summed to yield the cognitive/attention score. The externalizing score is computed by adding items 7, 8, 10, and 15. Finally, the internalizing score is created by summing items 9, 16, and 18. The reliability and change characteristics of scales composed of these items have been previously reported for school-aged children (e.g., Meier, 2004; Meier et al., 2008). Internal consistency has ranged from .72 to .86 for the total score.

PROPOSED MODEL TO EVALUATE SCHOOL-BASED THERAPEUTIC INTERVENTIONS

The IISR procedures coupled with the PBIS framework at the universal level provide a framework in which to assess the effectiveness of therapeutic interventions (i.e., tertiary-level supports). We propose a preliminary model using the BIMAS, a measure that provides empirical outcomes and is sensitive to change for evaluating school-based therapeutic interventions. The data collected from the BIMAS affords districts and mental health agencies the opportunity to evaluate therapeutic interventions provided in schools. Using the graphs suggested with a universal system of PBIS as a guide, we offer a similar approach for school-based mental health programs by considering the overall outcomes for individual students, whether these outcomes were different among types of therapies or particular diagnoses, if particular providers were more effective with students, and finally if the length of treatment affects outcomes. This translates into graphs reflecting the following information: (a) individual student, (b) treatment type, (c) diagnosis, (d) provider, and (e) treatment length (i.e., number of sessions). See Table 1.

For each of these categories, data can be evaluated by examining BIMAS ratings from clinicians, teachers, and/or parents. Average total scores would provide the first layer of analysis, and depending on the needs and desired outcomes of a district, these scores could be examined across clinician, teacher, and/or parent raters. If needed, index scores could be examined to provide a more refined analysis.

With data collected from a change-sensitive measure, a consistent set of graphs can be generated to aid in data-based decision-making practices. As similar with SWPBS, data to be graphed would include information assessing change by student, setting, intervention or treatment, and time. Specific goals were to develop a data-based decision-making model for teams at a school and district level reviewing tertiary level intervention programming (a) to assess the treatment gains made by individual children and youth receiving school-related services, (b) to assess the group gains of children and youth receiving a specified school-related treatment or a treatment by a specific provider, and (c) to provide a preliminary illustration of an evolving evaluation model that holds the promise for more fine-tuned assessment of school-based and school-linked tertiary-level services. Armed with these data, schools and districts have the tools to make data-based decisions not only for individual students but at a programmatic level as well. It should be noted, however, that given the sensitive nature of these data regarding students' social and emotional functioning, it is important to secure appropriate consents from parents for school-based teams to engage in planning in concert with clinicians.

Table 1. Comparison of SWPBS Data Graphs and Treatment Change Graphs

PBIS Graph	BIMAS Graph	Raters	Data
ODRs by student	Treatment change Individual clients	Clinician Teacher Parent	Mean total scores Mean index scores
ODRs by time	Change ratings by treatment	Clinician Teacher Parent	Mean total scores Mean index scores
ODRs by behavior	Change ratings by diagnosis	Clinician Teacher Parent	Mean total scores Mean index scores
ODRs by school location	Change ratings by site/provider	Clinician Teacher Parent	Mean total scores Mean index scores
ODRs by day/month	Change ratings by number of sessions	Clinician Teacher Parent	Mean total scores Mean index scores

## ILLUSTRATION OF A MODEL TO EVALUATE SCHOOL-BASED THERAPEUTIC INTERVENTIONS

To illustrate the approach, we present an example with actual data collected from elementary schools. Eleven students, ranging in age from 6 years 9 months to 11 years 2 months, who were receiving mental health services from a county mental health agency served as participants in the evaluation. Students were enrolled in one of five elementary schools with grades kindergarten through fifth in a small urban school district in the northeastern United States. Students were referred for services by teachers and/or parents/guardians. All mental health services were provided by therapists employed by the county and delivered in the student's home school during the school day or in school-linked neighborhood mental health clinics. The services were designed, delivered, and supervised by the therapist and the county agency.

For each student, clinicians were asked to collect demographic information (e.g., gender, age, date of birth) as well as assign a *Diagnostic and Statistical Manual of Mental Disorders* (4th edition) diagnosis and a global assessment of functioning (GAF) score at intake and discharge. During the course of treatment, clinicians recorded the type of treatment provided, therapeutic model utilized, length of treatment in weeks, number of meetings with school representatives, list of referrals for services, and clinical impressions. Treatment types included individual, group, family, collateral, informal consult, psychiatrist, or other. In terms of therapeutic models, 10 models were listed for endorsement of one or more by the clinician. Based on the endorsements of models, these were then classified as mixed or cognitive–behavioral strategies. The mixed model consisted of a combination of solution-focused, play-, cognitive-, educational-, and strength-based therapies. Clinicians were requested to rate participants on the BIMAS at intake and after mental health services were

terminated. Although parent information was requested, only three parents provided information, and this information was either provided at the first or final data collection point but not both, precluding the use of parent data in any further analysis. Although more common for data to be collected periodically throughout treatment within multitiered frameworks, collecting pre–post data improved substantially the availability of outcome data compared to standard practice. Notwithstanding, it is important that multiple types of assessment data are collected with students needing tertiary-level supports in accordance with the target behavioral concerns as well as data collected school-wide.

To illustrate this model, we will examine BIMAS total scores based primarily on clinician ratings, but we will also demonstrate how other types of information suggested could provide additional details. Examination of the first graph (see Figure 1) suggests that compared to total scores at intake, 6 of the 11 students exhibited improvement at discharge. That is, lower scores were obtained at discharge compared to intake. For the purposes of making specific decisions regarding individual students, the next step might be to look at the index scores for each of the five students whose symptomatology increased at discharge. For example, for student 9, adaptive and cognitive index scores were higher (18 and 8, respectively) than intake (13 and 7, respectively), and the same score was obtained for the internalizing domain (4). However, discharge scores (5) suggested improvements compared to intake (7) for the externalizing domain, which might be expected given that this student was referred for externalizing problems (see Table 1). Change in only those items most relevant to the treatment goals, therefore focusing on individual item analysis, is expected according to the IISR approach (Meier, 1997, 1998). In addition, given the relatively stable nature of many mental health problems (Hops et al., 1988; Reid, 1993), even small changes in item scores can suggest clinically meaningful improvements. Perhaps most important is that these changes are observed across settings within which the problem behavior(s) are occurring, highlighting the importance of monitoring across settings.

Figure 1. Average BIMAS total scores for individual students at intake and discharge.
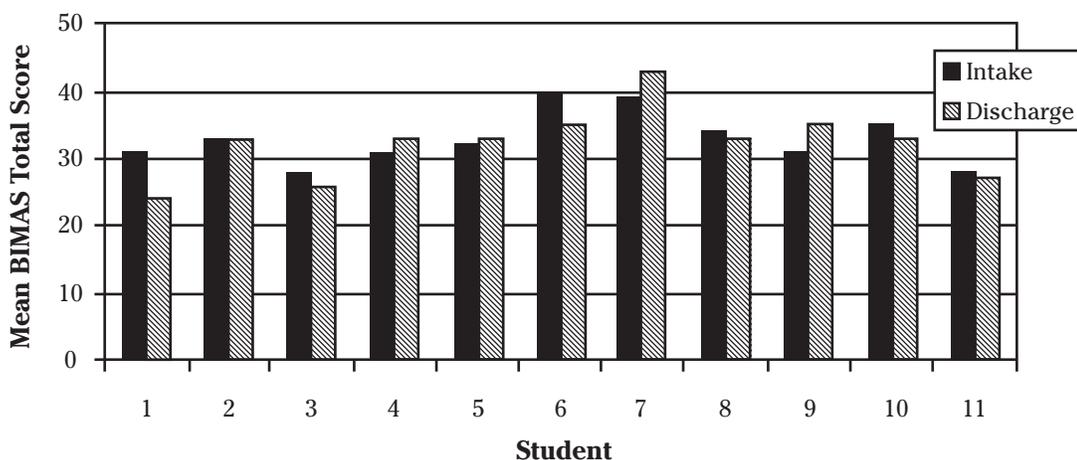
In Figure 2, the mean total score by type of treatment provided is depicted by clinician and teacher ratings at intake and discharge. Treatment modalities included cognitive–behavioral therapy or a mixed model of treatment. The figure suggests, according to clinician ratings, that on average students whose clinicians used the mixed approach to therapy made slight improvements whereas those who were treated using a cognitive–behavioral model did not. Because school personnel may also be interested in reviewing change according to teacher ratings to determine whether effects generalized beyond the clinician's office, teacher ratings are also included in Figure 2 for analysis. Unfortunately, in this case, teachers reported students' symptoms as similar as or worse than at the start of therapy. Although not available for these students, parent ratings would also be useful to further analyze the generalized effects of school-based mental health services.

Figure 2. Average BIMAS total scores by treatment type for clinician and teacher ratings.
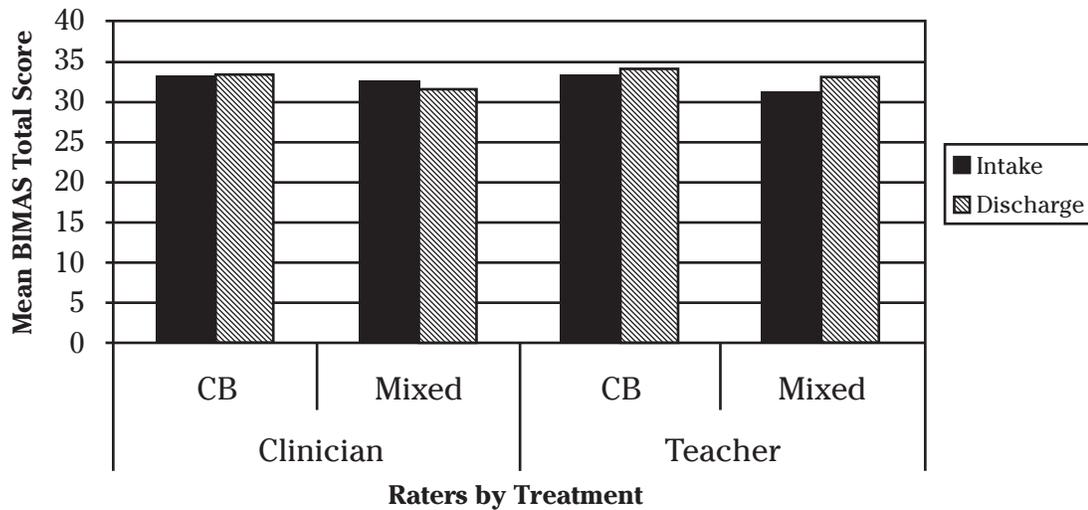

Figure 3 displays mean total score by diagnosis. Students were grouped into one of three categories (i.e., externalizing, internalizing, or comorbid), depending upon the presenting problem. Data are presented across the three diagnostic categories. These data suggest that children referred for internalizing and comorbid problems benefited from treatment, but little change was observed for children referred for externalizing problems.

## Figure 3. Average BIMAS total scores by diagnosis.
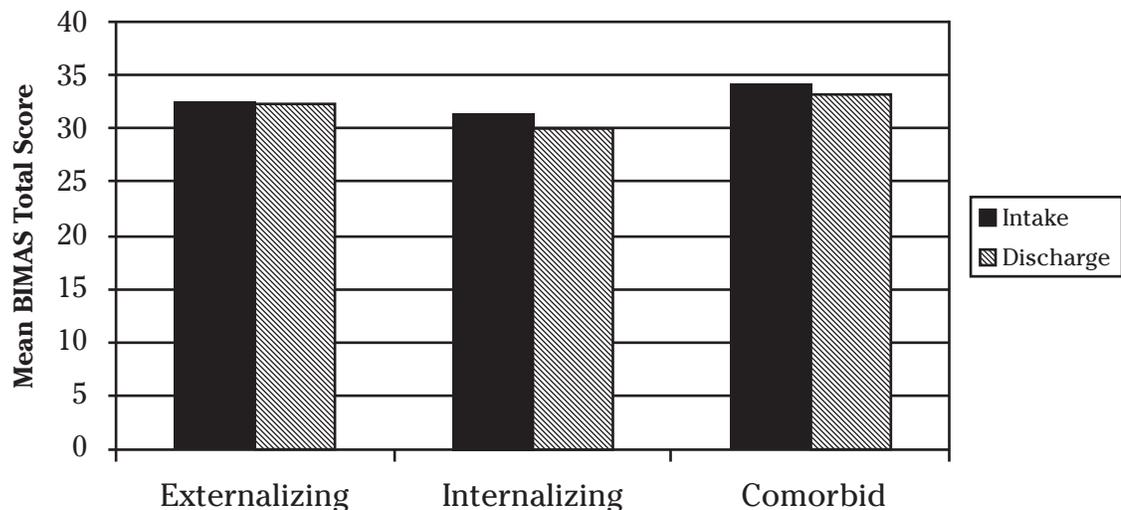
Mean total score by site/provider (i.e., clinician) at intake and discharge is displayed in Figure 4 by the clinician. Half of the providers rated changes in the desired direction across their caseload from intake to discharge. Investigating this issue further reveals that the providers who rated a decrease in symptomatology provided a mixed model of therapy. Similarly, this information could be put into context by looking at Figures 2 and 3. Figure 5 depicts the mean total score from intake and discharge by treatment length (i.e., the number of sessions attended; 0–10 sessions, 11–20 sessions, and 21–30 sessions), as rated by the clinician. Benefits of treatment were most notable when 10 or fewer sessions were provided, with more than 20 sessions also offering some positive change.

Figure 4. Average BIMAS total scores by mental health provider.
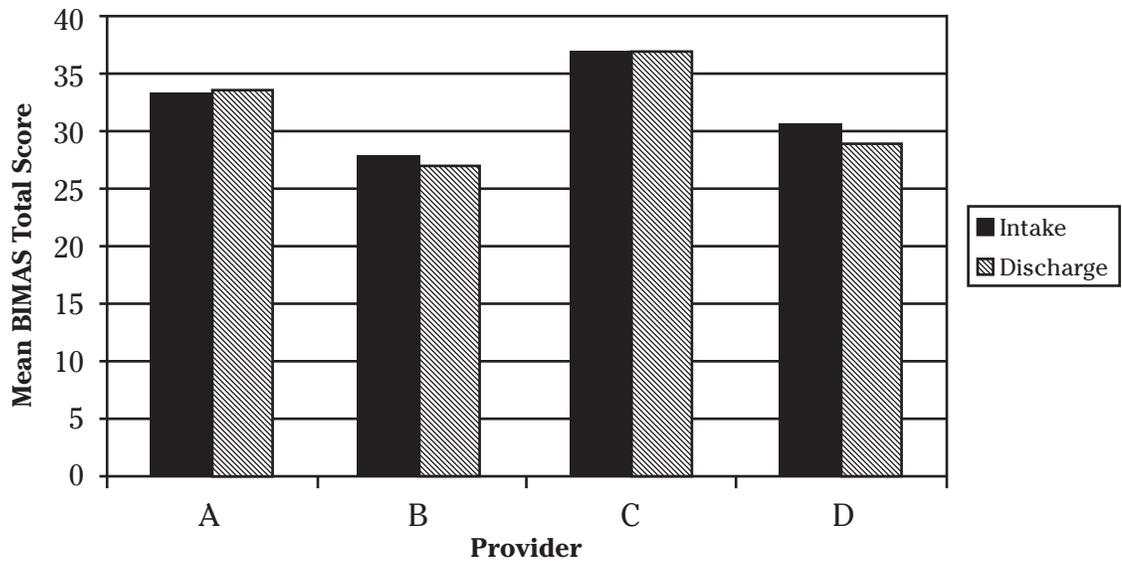
Figure 5. Average BIMAS total scores by treatment length.
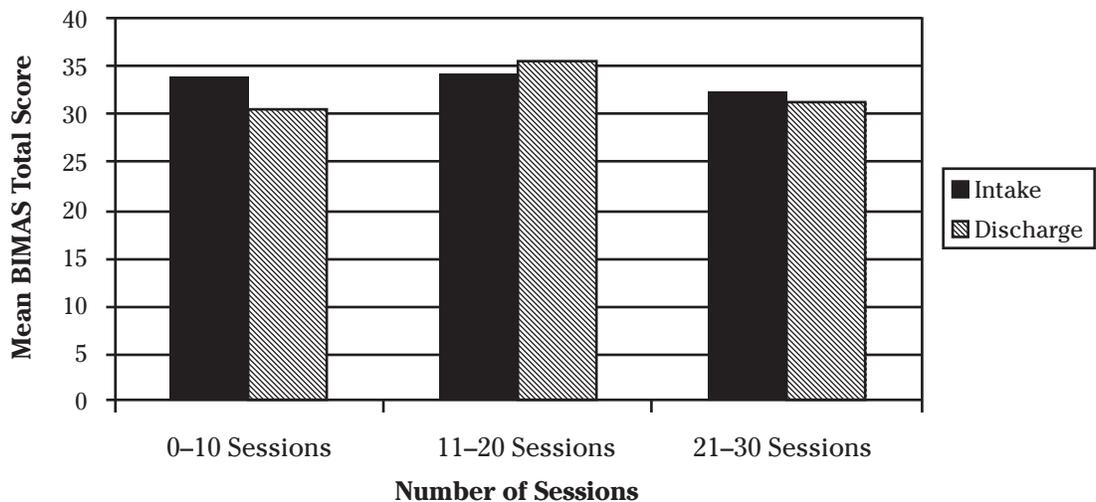
Overall in our sample, clinicians and teachers differed in their ratings on all levels. Clinicians rated improvements in functioning (i.e., decrease in total score) for students who participated in 0–10 treatment sessions, presented with internalizing concerns, and were treated with a mixed model of treatment. In contrast, teachers rated the treatment as not effective, resulting in an increase in total scores across all modes of analysis. Multitiered assessment and evaluation of treatment outcome is a critical component of wraparound services provided to students who need the most intense level of services (Scott & Eber, 2003). Therefore, evaluating treatment effects across stakeholders (e.g., parents, teachers, and clinicians) is important for schools to consider when monitoring and evaluating treatment plans. If, as in this case, data demonstrate improvements in one area (e.g., clinician ratings) but not another, then the focus of treatment can be modified to enhance collaboration across settings and/or target the elements of the environment in most need of additional support, which in this case was the classroom.

CONCLUSION

In order to address the number of children who experience emotional–behavioral difficulties and those who exhibit serious and chronic problems, recommendations have been made to organize schools around a three-tiered approach to school discipline where intervention is early and timely for those students in need of additional supports (Walker et al., 1996). At the tertiary level where children present with the most significant problems and often have the most risk factors, FBA-derived behavior intervention plans and wraparound services have been described as critical intervention components (Scott & Eber, 2003). Although measures have been recommended for use with respect to behavior intervention plans (e.g., direct measurement of behavior), schools also need to measure the outcomes for mental health services (Nabors, Wiest, Tashman, & Meyers, 1999). Traditionally, attempts to measure outcomes in mental health have relied on trait-based measures. Because these measures have been designed primarily for the purpose of diagnosis and emphasize stability over time, these measures may not be the best choice for evaluating intervention outcomes. Therefore, inclusion of measures that are sensitive to state-based changes may be important (Meier, 1997). Moreover, the impact of mental health services as seen by school professionals needs to be considered as part of the evaluation process and examined across individuals who live or work with students receiving services, in addition to the clinician. The purpose of this paper was to offer a model for assessing school-based mental health services employing a change-sensitive measure as evaluated by clinicians, teachers, and parents.

The five graphs generated provide information that can be used in a multiassessment format to examine effects of one aspect of wraparound services: mental health counseling. Recommendations for these particular graphs were intended to parallel the five primary graphs used to evaluate ODR data at the universal level of SWPBS. In this particular example, the BIMAS was used to examine mental health outcomes for 11 children presenting with a variety of significant problems. These data, as an example, can provide a district-based team with the necessary evaluative measures to make future decisions for the program (e.g., should the program continue in its current form, change the treatment modality, or discontinue). These types of decisions are significant at a program level and should only be made with data that are sensitive to student outcomes. At a local level, these data can also inform mental health providers and school-based tertiary teams that are charged with formatively evaluating the intensive interventions currently in place for students. In the current example, mental health providers may consider collaborating with teachers and other support staff to ensure that treatment is extended to the most problematic settings (e.g., lunchroom, playground) for these children. Clinicians, half of whom rated their own clients as exhibiting little change, could use this information to make data-based decisions regarding therapeutic treatment by periodically completing BIMAS ratings (Gray & Lambert, 2001; Harmon et al., 2005). That is, if students are not exhibiting as much progress as clinicians are expecting, types or approaches to counseling could be reconsidered during the course of treatment.

We also illustrated that more detailed evaluation of individual student data can be identified by looking at index scores. In this way, even if students are not exhibiting global changes as represented by total scores, growth may have occurred for specific target areas. Moreover, although students in this sample were not enrolled in schools implementing PBIS, this type of approach to mental health services could certainly be integrated within a PBIS framework. Integration of mental health and education, and in particular PBIS, has some preliminary evidence with PBIS providing an avenue for promotion of prevention across all tiers of the system (Kutash et al., 2006; Sprague, Nishioka, & Smith, 2007).

Data-based decision making is a fundamental practice for evaluating any intervention put in place. For students with chronic, serious problem behaviors best served through tertiary-level interventions, this practice is even more important, as the cost and resources required for multicomponent intervention implementation increase exponentially (Scott & Eber, 2003; Walker et al., 1996). Formative evaluation of tertiary-level services must be efficient and sensitive to change to inform a flexible model of service delivery as needs change (Eber et al., 2002). Paralleling the PBIS graphs of ODRs generated within a universal system, the BIMAS can produce data of similar scope and utility for tertiary interventions and yet still maintain some

level of feasibility. The appeal of ODRs as a formative measure is its availability as an outcome, which is particularly important at a universal level where processes are assessed and evaluated school-wide (Sugai & Horner, 2002). As schools move into individual systems of student support (i.e., secondary and tertiary levels) where the intensity of interventions must increase to match the intensity of student needs, the process of data collection must also increase in intensity (Walker et al., 1996). Feasibility of data collection is still a focus at the intensive levels of intervention. Given the emphasis on change sensitivity during its development, the multiple rating formats available, the efficiency of completion, and the ability to assess increases in adaptive functioning in addition to decreases in presented symptoms, we propose the BIMAS as one tool to consider for use in assisting the evaluation of school-based or school-linked interventions coordinated by tertiary teams.

Viewing PBIS under the lens of a school-based mental health model, PBIS has the potential to address students with significant behavioral concerns who have traditionally been served through a nonschool-based mental health system. In addition to the challenging task of design and fidelity of implementation for this level of intervention within the educational setting is data-based decision making for multicomponent interventions. We must look to methods that are sensitive, efficient, and technically adequate to evaluate effectiveness at an individual student and systemic level. As school psychologists expand their role to include prevention and intervention initiatives at all three tiers of service delivery, knowledge of change-sensitive measures is ever more important.

Limitations of Model Illustration Data

The model was illustrated with data that presented some limitations. First, the access to parent ratings was extremely limited, thus impeding the illustration of the model beyond clinician and teacher ratings. In future applications of a model, diligent efforts must be made to secure ratings from all three informants. Second, the data did not show significant changes in outcomes for the students. While this is disappointing, the intention was to demonstrate how the data could be utilized by schools. In this regard, lack of progress for students in school-based therapeutic interventions should raise a flag to local school and district teams on the potential effectiveness of the interventions. Actions may be taken by schools and districts to address the integrity of implementation, the appropriateness of the treatment for the student needs, as well as the evidence base of the treatment modality.

Implications for Practice

As schools move to three-tiered approaches, broad implications exist for school psychologists in practice. With expertise in data-based decision making, academic and behavioral interventions, systems reform, and assessment techniques, school psychologists can serve a pivotal consulting role on school teams addressing multiple tiers. The model proposed herein is one example where a school psychologist can assist with assessment and evaluation of a tertiary-level support. The BIMAS could be useful as one component of a multiassessment and evaluation method recommended for tertiary-level supports and could provide perspectives from parents, teachers, and the primary clinician. Consistent with multitiered models, this measure could be utilized three times yearly with students receiving the most intensive services.

Future Directions

This paper represents an example of examining data for the purpose of evaluating school-based mental health services with the goal of improving treatment and outcomes for children who require this support. These data represent the evaluation of treatment process (e.g., treatment type, provider) and outcome (e.g., total and subscale scores across raters) components of recommended quality assurance activities (Donabedian, 1980; Nabors et al., 1999) for school-based mental health services. Certainly, given the small number of students and clinicians providing therapeutic intervention in this particular example, it is important for this model to be applied to larger school-based mental health systems to further evaluate its

utility. It is also important to recognize that children in this sample were referred for services because they had particularly challenging or complex problems, which consequently may be reflected in the modest improvements. In addition, the sample reported here reflects only those students served for which data were collected. In practice, missing data would significantly limit the usefulness of these results, which emphasizes the necessity of allocation of time and resources to collect pre–post data from multiple informants. It is also important to note that these students were not receiving additional services, as would be reflective of a wraparound model. Similarly, investigation of other outcomes may be useful. The social validity of this model for teachers, clinicians, and administrators may need to be investigated in the future to evaluate whether its intended purpose is helpful. That is, student, parent, and teacher perceptions of successful treatment would also be important variables. Finally, whether evaluating these data result in effective changes in the mental health of school children should be considered in the future.

REFERENCES

Atkins, M. S., Graczyk, P. A., Frazier, S. L., & Abdul-Adil, J. (2003). Toward a new model for promoting urban children's mental health: Accessible, effective, and sustainable school-based mental health services. *School Psychology Review, 32*, 503–514.

Chafouleas, S., Riley-Tillman, T. C., & Sugai, G. (2007). *School-based behavioral assessment: Informing intervention and instruction.* New York: Guilford Press.

Clonan, S. C., McDougal, J. L., Clark, K., & Davison, S. (2007). Using data to improve student behavior: Data-based decision making and the positive behavioral interventions and supports project. *Psychology in the Schools, 44*, 19–28.

Donabedian, A. (1980). *The definitions of quality and approaches to its assessment.* Ann Arbor, MI: Health Administration Press.

Eber, L., Sugai, G., Smith, C. R., & Scott, T. M. (2002). Wraparound and positive behavioral intervention and supports in the schools. *Journal of Emotional and Behavioral Disorders, 10*, 171–180.

Gray, G. V., & Lambert, M. J. (2001). Feedback: A key to improving therapy outcomes. *Behavioral Healthcare Tomorrow, 10*(5), 25–45.

Harmon, C., Hawkins, E. J., Lambert, M. J., Slade, K., & Whipple, J. (2005). Improving outcomes for poorly responding clients: The use of clinical support tools and feedback to clients. *Journal of Clinical Psychology, 61*, 175–185.

Hops, H., Walker H. M., & Greenwood, C. R. (1988). *Procedures for establishing effective relationship skills.* Seattle, WA: Educational Achievement Systems.

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2005). School-wide positive behavior support. In L. M. Bambara & L. Kern (Eds.), *Individualized supports for students with problem behaviors* (pp. 359–390). New York: Guilford Press.

Kauffman, J. M. (1997). *Characteristics of behavior disorders of children and youth* (6th ed.). Columbus, OH: Merrill.

Kutash, K., Duchnowski, A. J., & Lynn, N. (2006). *School-based mental health: An empirical guide for decision-makers.* Tampa, FL: The Louis de la Parte Florida Mental Health Institute, University of South Florida.

Lambert, M. J. (1994). Use of psychological tests for outcome assessment. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 75–97). Hillsdale, NJ: Erlbaum.

Meier, S. (1997). Nomothetic item selection rules for tests of psychological interventions. *Psychotherapy Research, 7*, 419–427.

Meier, S. (1998). Evaluating change-based item selection rules. *Measurement and Evaluation in Counseling and Development, 31*, 15–27.

Meier, S. (2000). Treatment sensitivity of the PE form of the Social Skills Rating Scales: Implications for test construction procedures. *Measurement and Evaluation in Counseling and Development, 33*, 144–156.

Meier, S. (2004). Improving design sensitivity through intervention-sensitive measures. *American Journal of Evaluation, 25*, 321–334.

Meier, S., McDougal, J. L., & Bardos, A. (2008). Development of a change-sensitive outcome measure for children. *Canadian Journal of School Psychology, 23*, 148-160.

Morrison, G. M., & Skiba, R. (2001). Predicting violence from school misbehavior: Promises and perils. *Psychology in the Schools, 2*, 173–184.

Nabors, L. A., Weist, M. D., Tashman, N. A., & Myers, C. P. (1999). Quality assurance and school-based mental health services. *Psychology in the Schools, 36*, 485–509.

Nelson, J. R., Benner, G. J., Reid, R. C., Epstein, M. H., & Currin, D. (2002). The convergent validity of office discipline referrals with the CBCL-TRF. *Journal of Emotional and Behavioral Disorders, 10*, 181–188.

Nelson, J. R., Martella, R. M., & Marchand-Martella, N. (2002). Maximizing student learning: The effects of a comprehensive school-based program for preventing problem behaviors. *Journal of Emotional and Behavioral Disorders, 10*, 136–148.

Netzel, D. M., & Eber, L. (2003). Shifting from reactive to proactive discipline in an urban school district: A change of focus through PBIS implementation. *Journal of Positive Behavior Interventions, 5*, 71–80.

Office of Special Education Programs. (2004). *School-wide positive behavior support implementers' blueprint and self assessment.* Retrieved October 15, 2006, from www.pbis.org

Pelham, W. E., & Massetti, G. M. (2003). Commentary: Expanding on effectiveness. *School Psychology Review, 32*, 536–540.

Reid, J. B. (1993). Prevention of conduct disorder before and after school entry: Relating interventions to developmental findings. *Development and Psychopathology, 5*, 311–319.

Scott, T. M., & Eber, L. (2003). Functional assessment and wraparound as systemic school processes: Primary, secondary, and tertiary systems examples. *Journal of Positive Behavior Interventions, 5*, 131–143.

Sprague, J., Nishioka, V., & Smith, S. G. (2007). Safe schools, positive behavior supports, and mental health supports: Lessons learning from three Safe Schools/Healthy Students communities. *Journal of School Violence, 6*, 93–115.

Stiffman, A. R., Orme, J. G., Evans, D. A., Feldman, R. A., & Keeney, P. A. (1984). A brief measure of children's behavior problems: The Behavior Rating Index for Children. *Measurement and Evaluation in Counseling and Development, 16*, 83–90.

Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy, 24*, 23–50.

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 2*, 131–143.

Tobin, T., & Sugai, G. (1999). Using sixth-grade school records to predict violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders, 7*, 40–53.

Tobin, T., Sugai, G., & Colvin, G. (2000, May). Using discipline referrals to make decisions. *NASSP Bulletin, 84*(616), 106–117.

U.S. Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General*. Rockville, MD: Author.

U.S. Department of Health and Human Services. (2000). *Report of the Surgeon General's Conference on Children's Mental Health: A national action agenda*. Rockville, MD: Author.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., et al. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4*, 194–209.

Wright, J. A., & Dusek, J. B. (1998). Compiling base rates for disruptive behaviors from student discipline referral data. *School Psychology Review, 27*, 147–183.