

Running Head: Eco-Behavioral Intervention

Eco-Behavioral Assessment and Intervention for Culturally Diverse At-Risk Students

George H. Noell

Louisiana State University

Donna N. Gilbertson

Utah State University

Amanda M. VanDerHeyden

Vail School District

Tucson, Arizona

Joseph C. Witt

Louisiana State University

Submitted Draft for the Handbook of Multicultural School Psychology

Address correspondence to George Noell, Department of Psychology, 236 Audubon Hall,

Louisiana State University, Baton Rouge, LA 70803-5501. Phone: 225-578-4119.

Fax: 225-578-4125. E-mail: gnoell@lsu.edu.

Eco-Behavioral Assessment and Intervention for Culturally Diverse At-Risk Students

Schools are challenged to provide diverse students with experiences that result in learning. This heterogeneity of students' needs creates substantial challenges for educators. For example, students of varying linguistic backgrounds and degrees of facility in their first language will vary in the extent to which they will need direct instruction in phonemic awareness in English to become literate readers of English (Quiroga, Lemos-Britton, Mostafapour, Abbott, & Beringer, 2002). In addition to instructional issues, cultural diversity can create challenges for educators in creating learning environments that are conducive to learning, consonant with schools' values, and acceptable to parents. For example, religious and cultural differences in gender roles can create the context for conflict between families and schools over school activities and culture.

Within the context of diverse cultural and ethnic traditions, students exhibit substantial individual differences in their educational needs. To state the obvious, a student who is blind will have different instructional needs than a student who can see. Additionally, individual students who are blind have varied needs based on their learning experiences and other characteristics. These differences may be similar across cultural groups or they may interact with the specific traditions of the cultural and ethnic groups with which the student identifies. For example, Bau (1999) describes how differing cultural and family traditions can influence the provision of services to persons who are blind. In particular Bau describes the importance of considering linguistic barriers for non-native English speakers, the potential variability in goals for services, and being respectful of cultural traditions relevant to status within the family when providing services.

The example of a student who is blind and also happens to be a non-native speaker of

English provides a relatively unambiguous example of the potential for interaction between a student's cultural heritage and educational needs. However, the degree to which at-risk or disabled students' special needs are unambiguously evident varies widely. When the developmental challenge confronting a student is a reliably diagnosed neurogenic condition such as cerebral palsy, the student's needs for special supports in the domain of motor development is likely to be unambiguous. Additionally, the independence of the student's disabling condition and his or her cultural heritage or identity is clear.

However, students whose learning challenges are the result of reliably diagnosable genetic, sensory, and neurological conditions are the minority of students exhibiting special needs in schools (Fletcher et al., 2002). For the majority of students who do not learn well, no clear underlying bio-genetic disorder is evident that explains their specific difficulty with learning at school. These students have been described as difficult-to-teach (Bahr, Fuchs, Stecker, & Fuchs, 1991) or slow learners, and may also include students identified as learning disabled, mildly mentally disabled, or behavior disordered. Although numerous genetic or neurological abnormalities have been hypothesized or assumed to underlie learning disabilities (see for example Rourke & Conway, 1997), these abnormalities have not been demonstrated to have diagnostic utility within current diagnostic systems. For students who are described as at-risk, poorly achieving, or mildly disabled, their group membership is based on their behavior or symptoms. These symptoms are the failure to learn and behave in the manner that educators expect.

The identification of students with special needs based on their overt behavior creates additional challenges. Students' learning is a complex, multiply determined phenomenon. In a domain such as reading, variables associated with socio-economic status, culture, and individual

differences all contribute to reading success. For example, variables such as exposure to literacy in the home, exposure to standard English phonemics, and/or general intelligence can all contribute to literacy success (McGrew, Flanagan, Keith, & Vanderwood, 1997; Sonnenshein & Munsterman, 2002). Although social class, cultural, and individual difference variables will all contribute to varying degrees to all students' success in becoming literate, the extent to which any one or all of these factors is causal for a particular at-risk student may be difficult to determine or unknowable.

The ambiguity of differentiating between problems resulting from learning experiences versus a disabling condition is further complicated by the reality that for dichotomous decisions (i.e., disabled versus not) the reliability of decision-making is poorest for the population of students near the decision point boundary (Nunnally, 1978). When student behavior is near the decision point, it is difficult to consistently reach agreement on whether behavior constitutes a problem or is in the lower end of the acceptable range. Based on the distribution of skills, behaviors, and abilities in the population, the most common decisions will fall relatively near the boundary (e.g., mild disabilities such as learning disabilities). The reliability of decisions near the boundary and the recognition of the multiple causes of behavior creates a challenging context for making decisions. It may be unclear if a student's mathematics skills are weak enough to be a significant concern. It may also be unclear whether those weak skills are the result of disabling condition that is assumed to be inherent to the student, that student's learning experiences, or an interaction of the two. To further complicate matters, these difficult decisions occur within school-based teams consisting of teachers, administrators, and specialists (e.g., school psychologists) who may focus on the teacher's complaint rather than the child's characteristics or the legalities of classification and placement. Perhaps this helps to explain the finding in one

field-based examination that as many as 50% of the studied children who were placed in special education did not meet the eligibility criteria for their identified disability (MacMillan, Gresham, & Bocian, 1998).

The intersection of cultural diversity, individual diversity, and the identification of students as disabled has been one of the central nodes of conflict and litigation in the United States over the last several decades (see Reschly & Bersoff, 1999). The extent to which assessment and decision-making are fair and unbiased has been at the center of this controversy. Bias in psychological assessment has been defined as the extent to which irrelevant variance is present within obtained scores and that variance is differentially associated with specific groups (Messick, 1995). In schools the critical issue has been the extent to which racial or cultural group membership has influenced access to services and placement in segregated programs. Overrepresentation in special programs does not demonstrate bias per se (Cole, 1981). If a particular group has a higher incidence of need and the assessment process selects them based on that need, then the process would reflect the underlying reality rather than being biased. On the other hand, if the assessment process over-selects or under-selects students as needing specialized services as a result of variability associated with their cultural group (i.e., systematic error raising or lowering scores for groups) then the assessment and decision process would be biased.

The assessment processes used by schools to identify students who need additional support or who are disabled has substantial consequences for those students. In some instances concerns have been raised that procedural or instrument bias has deprived students of needed services. More commonly, potential bias based upon cultural group membership has been regarded as suspect due to concerns about the segregation that is commonly attendant to

classification of a student as disabled and the generally poor efficacy data for special education (Artiles & Trent, 1994; Kavale & Forness, 1999). The societal tensions between access to services, the stigma associated with labels, the dominance of segregated services, and concerns about the efficacy of those services have maintained the controversy surrounding assessment and special services in schools. This controversy over the extent to which educational referral, assessment, and placement decisions are biased toward placing minority students in programs for the mildly disabled has also been the focus of considerable litigation (Reschly, Kicklighter, & McKee, 1988). Reasonably, the question has been raised as to why students who are members of some minority groups have been overrepresented in largely segregated programs in some jurisdictions. This question is particularly salient in light of the generally poor outcome data for special education in particular.

The extent to which assessment, decision-making, and educational practices are culturally fair is a related, but distinct issue. Determination of cultural fairness is more intractable because cultural fairness lacks an accepted professional definition with specific means of evaluation. The fairness of an assessment, decision-making, and service delivery process includes diverse issues such as the provision of due process, equivalence of implementation across groups, achievement of similar positive outcomes, and consideration of differences in goals or values across individuals and cultural groups. Whereas the extent to which an assessment device or process is biased can be tested against a specific model of test bias such as the regression model (Cole, 1981), the extent to which a practice is fair is a multidimensional evaluation of nonequivalent dimensions. For example, educational assessment may include due process and be respectful of different cultural practices, while at the same time leading to disproportionate negative outcomes for members of specific demographic groups. In this example, the practice may be culturally fair

in some ways and unfair in others.

Concern for disparate outcomes for groups is the bedrock for concerns regarding test bias and cultural fairness in assessment, decision-making, and service delivery. Assessment practices can only be biased or culturally unfair to the extent that they result in different outcomes across groups. However, the recognition of cultural groups is a social and political process in which the boundaries are ambiguous and whose variability within groups may exceed the variability across groups. Afridi (2001) describes the diversity of people and subgroups within the American Muslim population as a case in point. Afridi argues that the monolithic picture of Muslims portrayed in the popular media obscures the enormous heterogeneity of Islamic Americans. American Muslims vary widely in religious beliefs, religious practices, family customs, educational aspirations, and the roles of women within families (Afridi, 2001). The degree of variability may be similar to Americans independent of religious belief. Additionally, the accepted boundaries for group membership are not stable with the passage of time. Recognized groups merge over time to such a degree that they are no longer recognizable as distinct and new subgroups emerge (i.e., the emergence of African American Islam of conversion). The ambiguity of, changing nature of, and variability within cultural groups makes the development and maintenance of unbiased culturally fair practices a substantial and ongoing challenge.

Traditional nomothetically based practices have been intimately involved in the controversies over the identification of students who are members of minority groups as disabled due its fundamental emphasis on groups and its use in making diagnostic decisions in schools (Reschly & Bersoff, 1999). Scores are described as existing within a distribution drawn from a population as represented by a normative group that was sampled so as to be representative of that population. The population is a fundamental concept for nomothetic psychometrics.

Critical challenges for normative nomothetic comparisons are establishing the degree to which the normative sample is adequate, the degree to which the assessment instruments perform similarly across populations and sub-populations, and identification of the appropriate population.

The synergy that has evolved between societal concerns for cultural equity, nomothetic assessment traditions, and the challenge of diversity within schools appears to have created an intractable problem. Schools are confronted with students who exhibit substantial differences from their peers in their learning and behavior. Schools attempt to classify individual differences to identify the apparent problem, certify students' needs, and guide programming.

Unfortunately, substantial data exist demonstrating that the current categories for students with mild needs are not reliably distinguished and data demonstrating their instructional utility is lacking (Fletcher et al., 1998; Vellutino, Scanlon, & Lyon, 2000). To further confound the problem, the diagnostic categories are then layered across cultural categories creating a grid in which varying proportions of each cultural group are also disabled or not disabled. The crossing of cultural and individual difference categories naturally raises concerns regarding bias and the equity of the treatment of individuals within particular cells. The process of crossing individual difference groups with broad, permeable, overlapping, and evolving socio-cultural groupings sets the occasion such that a reasoned argument should always be available that some form of bias and inequity is present in schools. The multiple dimensions of individual differences and cultural variations seems to ensure that at any instance in time, some groups will be underserved, over represented, or afforded inadequate consideration of their needs.

Establishing need through Eco-Behavioral and Idiographic Assessment

Traditional nomothetic assessment emphasizes groups, differences between groups,

norms, and the creation of new groups based on test scores. The emphasis on groups with traditional nomothetic approaches to decision making may help perpetuate concerns regarding equity and bias rather than resolving them. In a context in which people are treated differently based on the groups they belong to, the natural conditions for substantive concerns about equity and bias seem to have been established. Interestingly, an alternative model for decision making within psychology and education has existed for many decades that de-emphasizes groups, populations, and norms. Idiographic assessment and intervention approaches have traditionally emphasized those characteristics that are common across humanity and how these characteristics are expressed in specific individuals.

Major theorists have advocated the idiographic study of human behavior, but have varied considerably in the methodologies they have recommended (i.e., Gordon Allport versus B. F. Skinner). The behavior analytic tradition appears to be a particularly promising approach for the incorporation of idiographic data into decision making about students exhibiting learning difficulties for several reasons. First, behavior analysis emphasizes a low level of inference in which the behaviors of interest are measured with little or no reference to unobserved constructs (Baer, Wolf, & Risley, 1968). Second, behavior analysis has an extensive research literature describing specific methods for collecting data and making decisions across a wide range of concerns evident in children. Third, behavior analysis emphasizes those principles of behavior that are common across humanity, while at the same time being specifically focused on the expression of that principle in an individual (Skinner, 1953). Finally, behavior analysis is specifically focused on changing behavior to socially significant degree to improve individuals' life and functioning (Wolf, 1978). This is consonant with schools' mission and the broader societal concerns.

The balance of this chapter describes a general model for assessment and intervention for students who exhibit problematic learning and behavior. The approach is described as eco-behavioral to connote the inclusion of both behavior analytic methodology and relevant assessment procedures such as criterion-referenced assessments that have been developed within education. The description will begin with a brief discussion of some fundamental elements of this approach and how this is related to student diversity. This is followed by a more detailed description of a specific eco-behavioral model that has been developed and field-tested based on the principles described herein. That model is the Screening to Enhance Equitable Placement (STEEP).

Elements of an Eco-Behavioral Approach to Diversity

In the foregoing discussion we have begun to build a case that assessment practices which rely on a nomothetic tradition, may, at the least, make school-based problem-solving for an *individual* child more difficult. This can occur because of a failure to take into consideration the *context* of the teaching and learning for the individual. Context matters. Consider, for example, a third grade classroom in the Hill District in Pittsburgh which as been described as one where the children are coming out of housing projects through the crack and gang infested neighborhoods. Such inner city schools often have the least qualified teachers because those who are certified and excellent often opt for less challenging environments in which to teach (Donovon & Cross, 2002). Many third grade children in this district are at-risk. The majority of children perform markedly below grade level on academic tasks. If any one child is referred and subjected to a nomothetic analysis, that child may exhibit a severe discrepancy between where they would be expected to be functioning and where they are functioning. This finding can be used to entitle the child to special education. However the problem is probably not

disability. The problem is not race. The problem is most likely that the child started school behind national averages in academic skills and has not been exposed to effective teaching that allow them to catch up. This problem would most likely be found by examining context. Specifically, if many children in a school are markedly deficient in reading, then we must begin to suspect factors other than disability. Legally we are required to examine context and to rule out cultural or educational disadvantage. However, most school-based teams are ill-equipped to measure and operationalize this ecological processes. An eco-behavioral analysis of learning is particularly well suited to help teams to make better decisions about *what* the problem is and *how* to resolve the problem.

It can be argued that the central element of all professional decision-making is data and rules for evaluating that data (Haynes & O'Brien, 1999). Shifting what data are collected and how those data are evaluated may result in educational approaches for culturally diverse at-risk students that are less susceptible to bias and that may result in better outcomes for students. Specifically, by lessening the emphasis on comparisons to national normative samples for the purpose of assigning students to categories, the issues of bias and unfairness may be attenuated. An alternative focus on comparison to criterion based standards, local norms, and the assessment of response to intervention may provide a means for psychologists to provide support for students who are at-risk that may reduce the need to further segregate students into additional categories based on disability or severe problems.

Data Collection

Shifting from a nomothetic diagnostic approach to an idiographic intervention focused approach (eco-behavioral) requires a number of changes in current educational practices. The first change is the focus of assessment. Within an eco-behavioral approach the problematic

behavior is measured directly rather than sampling behaviors within standardized test batteries designed to make inferences to unobserved constructs. It is important to acknowledge that the degree to which a particular standardized test directly assesses a specific skill (e.g., decoding) versus a broad construct (e.g., global reading) varies. The critical issue for an eco-behavioral approach is that a concern in the domain of reading is a beginning point that can be followed up with more detailed assessment of specific reading behaviors.

An additional element of deciding what to assess is the selection of critical targets for screening and progress monitoring. The concept of keystone behaviors can be quite helpful in identifying target behaviors (Barnett, Bauer, Ehrhardt, Lentz, & Stollar, 1996). A keystone behavior is one which is necessary for success in a broad array of situations or which facilitates success in a number of socially important situations. Reading is an obvious example of a keystone behavior in schools and in the community. Even within reading there exists an enormous array of additional subordinate skills that may not be equally important (e.g., rhyming versus comprehending). An additional element of the eco-behavioral approach described herein is focusing on those keystones within academic domains that relate to multiple positive outcomes. For example, prior research has demonstrated that phonemic awareness is related to the development of decoding skills and that fluent decoding is related to important reading outcomes such as comprehension (Snow, Burns, & Griffin, 1999).

A final emphasis in the selection of targets for assessment is stressing socially valued behaviors. Rather than focusing primarily on arbitrary and disembedded tasks (e.g., the Block Design subtest of the Wechsler Scales), an eco-behavioral assessment will focus on those behaviors that are important to educational success in the natural context. This is not to indicate that contrived assessment activities are not useful for answering some specific questions. The

emphasis on assessment of important behaviors in the natural context is akin to the zeitgeist underlying authentic and portfolio assessment (Elliott, 1991; Roe & Vukelich, 1998). The primary difference is that in the tradition of behavior analysis, eco-behavioral assessment will emphasize assessment in an objective, reliable, and continuous manner.

Data Evaluation

Arguably, selecting useful and unbiased methods of evaluating data may be more challenging than deciding what to measure. Within nomothetic assessment, a student's performance or the difference between performances on two tests is compared to a norm. If that performance or difference is beyond some standard, the assignment of the student to a group may be supported. Typically this group is described as disabled. The criticisms of this approach's utility and bias are legion and beyond the scope of this chapter (see for example Fletcher et al., 2002; Gresham & Witt, 1997; Vellutino et al., 2000). Whatever the limitations of the nomothetic approach, it does provide an objective systematic means of identifying students with exceptional needs that has the potential for being reliably implemented (despite evidence that in practice it may not be, MacMillan et al., 1998).

A critical element of an eco-behavioral approach is that it seeks to maximize the extent to which judgments can be arrived at objectively and that those judgments are useful. In the context of an eco-behavioral assessment, useful can be taken to mean having utility in planning instruction and accurate in identifying students who need additional supports. To achieve these ends the combined use of idiographic, local normative, and criterion referenced comparisons applied to the same data are employed.

In the current context, an idiographic comparison connotes a comparison of the student to him or herself under differing conditions of data collection or across time. The intent of this

assessment is to identify those conditions under which the student's performance improves or the student demonstrates learning. For example, recent research has shown that exposure to a series of brief test conditions can help to differentiate students whose academic needs are primarily motivational versus those with primarily instructional needs (Noell, Freeland, Witt, & Gansle, 2001). The critical element of these idiographic comparisons is the identification of promising strategies for the promotion of learning and then the monitoring of progress when that strategy is implemented. The details of the devising of these idiographic comparisons are beyond the scope of this chapter. Interested readers might consult Wolery, Bailey, and Sugai (1988); Howell, Fox, and Morehead (1993); or Kelly, Reitman, and Noell (2002).

Although idiographic comparisons are the only type of comparison that can establish what is an effective teaching strategy for a student is, they do not permit decisions about need in and of themselves. Determination of need requires comparison of the student's behavior to some standard. The traditional route has been to rely on well-developed, psychometrically sound standardized tests. Unfortunately, these comparisons are affected by a number of variables beyond the student's abilities or needs. For example, standardized test comparisons cannot account for the varying degree to which the curriculum in which the student was instructed does or does not overlap with the test. The degree of overlap can substantively affect student performances (Bell, Lentz, & Graden, 1992). Similarly, differences in access to high quality instruction will affect performance, but are not captured by nationally normed tests.

Comparison of student performance with a local norm based upon a classroom, school building, or school district can create considerably greater homogeneity of instructional experiences, curricula, and cultural backgrounds within the norm group. A local norm can be used to create a locally meaningful comparison that addresses an important question: How does

the referred students' performance compare to other students who have had similar opportunities? This type of comparison can create a more meaningful comparison in that a large number of potential confounds are held constant. See Ikeda, Tilley, Stumme, Volmer, and Allison (1996) for a description of the implementation of a service model that incorporates local norms across a large number of schools and school districts.

Although local norms can permit apples to apples comparisons, they also foster additional challenges. They will be impractical to develop if the behaviors to be normed require lengthy assessment. This limitation can be overcome by focusing norming on a limited set of keystone behaviors or benchmarks and by using time-efficient inexpensive measures such as those devised within curriculum based measurement (Shinn, 1989). A second potential concern is the impact of extreme environments on local norms. For example, a third grade student in a dysfunctional school may know 80% of the single digit addition facts and yet be only slightly below the local norm. Addition facts are a basic building block skill that should be mastered with 100% accuracy and reasonable fluency before entering the third grade. In this case, an extremely low group norm can make a student's performance appear average, despite the reality that when it is compared to a reasoned criterion it is unacceptable. In a similar sense, a very high performing group can result in student performance above a reasonable criterion standard falling in the lower extreme of the local normative distribution.

Buttressing local normative comparisons with the addition of criterion-referenced comparisons provides a means of arriving at more reasonable decisions regarding which students exhibit a substantial educational need. For example, discovering that a referred fourth grade student reads 35 words per minute and is at fiftieth percentile in his class might suggest that no problem is evident in reading decoding. However, comparison of this reading fluency to

criterion standards (e.g., Fuchs & Deno, 1982; Howell, Fox, & Morehead, 1993) would suggest that the student reads at less than half the fluency that would be expected at this grade level.

This suggests the need for additional instruction or support of the referred student as well as the majority of his classmates.

The intent of the three points of comparison described above, idiographic, local normative, and criterion, is to provide educators information that is useful and provides reasonable protections to the diverse body of students for whom concerns will arise. The initial comparisons would typically be to local norms and criterion standards. Comparison to local norms provides a benchmark against students within the same school culture. It provides evidence as to whether the student's learning is discrepant from other students attending the same class, school, or school district. The comparison to a criterion standard in combination with the local normative comparison should also help to clarify whether problem is systemic, many or most students perform poorly, or whether it is a student-specific concern. The final protection, examining the students' learning when they are provided systematic instruction may provide the most important protection. It examines the possibility that the student will learn well if he or she is systematically instructed.

The goal of adopting an eco-behavioral approach to coping with individual differences in a multicultural context is to emphasize the collection of information that has utility in solving problems. Additionally, an eco-behavioral approach seeks to avoid those nomothetic constructs that may have inhibited educators' success in developing systems for coping with individual needs that are broadly regarded as unbiased, equitable, and effective. The extent to which this approach can be successful in developing unbiased, equitable, and effective systems is largely unknown. In practice this will largely depend on the ability of educators to evaluate data

objectively and act on those data. Previous research on the decision processes of educational professionals suggests that is a dimension of the process that will require scrutiny (MacMillan et al., 1998). The remainder of this chapter describes a specific operationalization of the eco-behavioral model for coping with individual needs in multicultural contexts that has been developed, researched, and field tested. The process to be described, termed Screening to Enhance Equitable Placement (STEEP), is not the only possible eco-behavioral approach to this problem; it is a concrete example that is being used in schools in several states.

Screening to Enhance Equitable Placement

In this section, we describe STEEP (Screening to Enhance Equitable Placement), which is a practical, evidenced-based model for improving academic achievement and reducing the need for special education. STEEP was derived from a foundation of direct measurement of meaningful behaviors and allows professionals to systematically consider variables that are directly relevant to a child's present functioning and which have direct implications for intervention. From a practical perspective, STEEP is less concerned with *traditional explanations of why* minority children perform in a particular way because such explanations are nominal and nomothetic. Nominal or naming explanations connote explanation and understanding by naming. Attention Deficit Hyperactivity Disorder is a name for a variety of behaviors that adults object to. The label does not help us understand how to treat those behaviors. A nomothetic explanation is one where a child is thought to be problematic or disabled because he or she differs from the norm. Often nomothetic and nominal thinking is combined such that a child who deviates from the norm is given a label or name.

Thinking about minority children has frequently derived from a nomothetic conception of humanity. For example, traditional conceptualizations of bias start with the observation of

African American children scoring lower on a certain test. Cultural or racial differences are offered as at least partial explanation. In reality, race, per se, explains very little about a particular child's performance. Consider the following statements about Brian:

- 1. Brian is African American and is a poor reader.**
- 2. Brian is in a low performing school where children have not been taught to read well. He was "tested" with two reading interventions. Intervention A produced slow growth but intervention B produced rapid growth when both interventions were used for 2 10-minute sessions.**

The first statement is nominal; it names which groups Brian is a member of. The latter statements describe how Brian responds to changes in his environment. Which of the above two statements would teachers working with Brian find most useful? We are not suggesting that race is unimportant. Studies that illuminate the interaction between race and learning may ultimately add to our understanding of how to maximize learning for more children. However, an analysis of the variables that are proximal to learning and which schools can change is likely to be immediately and consistently productive.

STEEP Overview

The guiding principles of STEEP are (a) behaviors should be directly measured, (b) procedures and explanations should be parsimonious, (c) efficiency and effectiveness of procedures should be measured and ensured, (d) assessment procedures must have treatment validity and be designed to assist professionals to remediate and improve the problems identified, and (e) data inform actions and are used to evaluate effectiveness in an ongoing fashion. The goal is to give practitioners direct methods to collect and use data to inform decision-making in the schools. Each activity of STEEP yields data. The data are available immediately and

organized in a way that meaningfully and simply communicates the relevant information to the school-based team and guides the action that follows. The basic components of STEEP are research-based and have been studied in a program evaluation in one school and yielded promising findings (VanDerHeyden, Witt, & Naquin, in press; VanDerHeyden & Witt, under review). In addition, they have undergone extensive field testing in schools in multiple states. A description of the basic components of STEEP follows.

Schoolwide (Universal) Screening. STEEP is a problem-solving model of assessment that uses school wide CBA and CBM screening to identify performance problems in schools. This means that all children in a school are screened in reading, writing, and math. Following school wide screening, problems can be categorized as a classwide (i.e., class median score falls below the instructional range described by Deno & Mirkin, 1977) or individual child problem (i.e., classwide median falls at or above the instructional range and individual child scores below the 16th percentile for his or her class and in the frustrational range described by Deno & Mirkin, 1977). Thus, two anchors are applied initially to define the problem. The local anchor is classwide performance and the broader anchor is instructional level performance indicative of functional competence. Figure 1 shows a graph of how the classwide assessment data are organized so that the teacher and the school-based team can review the performance of all children at a glance.

Performance/Skill Deficit Assessment. Once a classwide skill problem has been ruled out, children who perform below the 16th percentile for their class and fall below the instructional range participate in a brief assessment of the impact of powerful incentives upon performance (i.e., performance/skill deficit assessment). During the performance/skill deficit assessment, the consultant provides the student with a copy of the classwide academic assessment probe that had

been previously administered. Students are told that they can earn a reward of their choice from the treasure chest by “beating their previous score.” This score is written in the top left-hand corner of the student’s paper. Students are allowed to sample briefly the items in the treasure chest. The treasure chest is a small transparent box containing several small tangible items (e.g., pencils, balls, stickers, bracelets, coupons for free time). The probe is then re-administered using the directions described above. The performance/skill deficit assessment for math can be administered to groups of three to five students simultaneously, whereas the performance/skill deficit assessment of reading is administered individually in a quiet space on the school campus. This component requires no more than five minutes per assessment. Children whose performance improves to the instructional range to earn an incentive do not participate in further assessment. Children whose performance does not improve to the instructional range will participate in an individual intervention in their classroom. These procedures are based on previous research suggesting the treatment utility of this type of assessment (Noell et al., 2001; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997).

Intervention. Interventions are directly linked to the type of problem identified. The design of intervention is intended to maximize the treatment utility of assessment (see Hayes, Nelson, & Jarrett, 1987). When a classwide problem is identified, a *classwide intervention* is prescribed. To develop an appropriate classwide intervention, classwide probes are administered using adapted survey-level assessment procedures to identify the instructional level such that the class median falls within the instructional range. Classwide intervention consists of paired peer practice directed toward fluency-building (e.g., flashcards for math skill) followed by independent timed probes of that skill, followed by error correction. This type of intervention generally requires ten minutes daily. The classwide intervention is performed for five to ten

consecutive days to identify the children who continue to perform below the instructional standard and demonstrate poor growth. Those children who do not progress during classwide intervention continue to perform as an at-risk learner. These children are identified for additional intervention which is designed to more specifically address their needs.

The second type of intervention is the *individual* intervention. This type of intervention is appropriate in situations for those children exhibiting skill deficits in classes where the majority of the class is performing at or above the instructional range. Typically, daily intervention is performed by the classroom teacher (or teacher designee) in the regular classroom setting during the regular school day. In this stage, a standard evidence-based intervention that requires approximately ten minutes per day is used. Intervention can begin immediately following screening and is linked directly to the assessment data. The student's instructional level is determined by sampling backward through successively lower level materials until the student scores in the instructional range. Protocol-based interventions consist of four basic steps: modeling, guided practice with immediate error correction, independent timed practice with slightly delayed error correction, and the opportunity to earn a reward (from a treasure chest) for "beating the last highest score." Each component is programmed in the intervention based on previous studies supporting the effectiveness of each component (Barbetta, Heward, Bradley, & Miller, 1994; Bennett & Cavanaugh, 1998; Daly, Lentz, & Boyer, 1996; Daly & Martens, 1994; Daly, Witt, Martens, & Dool, 1997; Rhymer, Skinner, Henington, D'Reaux, & Sims, 1998; Skinner, Cooper, & Cole, 1997).

The interventions are protocol-based and designed to produce evidence (i.e., permanent products) that they occurred to allow for estimation of treatment integrity. The school-based specialist collects the data weekly, quantifying two critical variables: the degree to which the

intervention occurred correctly and the child's performance on a novel, instructional-level probe of the target skill and a novel, criterion-level probe of the target skill. The specialist enters the data into the database and graphing tools automatically generate graphs for the teacher, principal, and school-based specialist. If integrity problems occurred, then the specialist re-trains the teacher. This general process for monitoring and intervening upon treatment integrity is based on a series of studies examining means of monitoring and enhancing treatment integrity (Noell, Duhon, Gatti, & Connell, 2002; Noell et al., 1997; Witt, Noell, LaFleur, & Mortenson, 1997)

The purpose of this brief intervention is to answer the assessment question: *Does the referred child appear to learn as a normal learner?* Hence, the goal of the intervention initially is to answer an assessment question about the child's response to intervention. To answer this question, a minimum of 10-15 consecutive intervention sessions, conducted with integrity, are required. Intervention progress for a student (i.e., slope derived from ordinary least squares regression) is evaluated relative to data pertaining to "normal" progress as well as data pertaining to the progress of children receiving specialized intervention. Structured decision rules are applied to determine if the child is a "responder" or "nonresponder" to the intervention or whether additional data are needed. Data on nonresponders are made available to the school-based team to assist in determining whether or not a child should receive an eligibility evaluation.

The emphasis on using intervention data to answer an assessment question provides a dynamic assessment that can address criticisms that traditional assessments are content invalid, indirect, insensitive to change (Marston, 1989), and have failed to demonstrate social and treatment validity (Gresham & Witt, 1997; Macmann & Barnett, 1999). In contrast, progress monitoring with CBM can be conducted in the regular setting, allows for time-series

representation of student performance, and allows for an evaluation of classroom context (i.e., instructional variables). The effectiveness of the intervention can be measured in an ongoing fashion to enhance effectiveness and solve problems ineffectiveness so that instructional opportunity is not lost and child outcomes are improved (Shinn & Bamonto, 1998). When using CBM within a problem-solving framework (Good & Kaminski, 1996; Shinn, 1989), the problem is defined in terms of hypothesized causes of deficient performance. Instructional modifications are first attempted in the current instructional environment. Assessment procedures are then designed to confirm or disconfirm the hypothesized causes of the problem performance, and intervention activities are designed to improve student outcomes in measurable units. Thus, two criteria are critical in determining whether or not a student exhibits a learning problem that warrants further assessment for potential special education services (Fuchs & Fuchs, 1998). The student should score below a criterion level of performance (compared to national standards and compared to classmates) and the student should demonstrate a growth rate less than that of his or her classroom peers. The effectiveness of the problem-solving model depends upon how well the problem is solved (i.e., student performance), and thus, ensures measurement of treatment validity (Messick, 1995; see also Fuchs & Fuchs, 1998).

STEEP simplifies and specifies exactly which activities should be conducted at which time and how the resulting information should be used to accomplish accurate identification. It reduces the need for full and individual evaluations because it rules out children who are responsive to intervention and/or are similar to their peers in achievement. These “rule outs” are consistent with IDEA provisions that some children are not good candidates for placement because their problems are not the result of a disability but instead stem from factors such as educational disadvantage, motivational problems, or a lack of exposure to quality instruction in

general education. In order to obtain a relatively pure measure of child competence, many potentially “masking” variables must be controlled as part of the process. In the figure below, we attempt to illustrate some of these variables that impact intervention effectiveness and will need to be addressed in an effective response to intervention (RTI) model. (See Figure 2).

The accuracy of the first two stages of STEEP with 182 participants received the school wide CBM assessment as well as the skill/performance deficit assessment was evaluated in a controlled investigation (VanDerHeyden, et al., in press). In addition, participants received a single five-minute instructional session to clarify the task and model correct responding. With these three pieces of information included within the STEEP model, a comparison was made between STEEP and teacher referral, the most common means by which children come to the attention of the school based team. Positive predictive power (compared to a criterion that included full curriculum-based assessment combined with extensive individual intervention) for traditional teacher referral was .19 compared to .53 for STEEP. This means that only 19% of the children referred by their teacher were identified as exhibiting substantial problems in the extended assessment where as STEEP was in accord with the criterion 53% of the time. Negative predictive power was .95 for STEEP indicating that model correctly identified 95% of true negatives. When Iowa Test of Basic Skills was used as the criterion, negative predictive power was 1.0 (VanDerHeyden et al., in press). STEEP also identified children who were overlooked by teachers (negative predictive power for teachers was .89). Finally, teacher referral was markedly affected by context in that predictive power varied dramatically across high- and low-achieving classrooms for teacher referral, whereas STEEP predictive power remained stable (VanDerHeyden & Witt, 2003). When applying STEEP, approximately 6% of the sample proceeded through to individual intervention. Approximately 15-16% of children were identified

through the school wide screening. Approximately 11% of the total were found to exhibit a skill deficit that warranted individual intervention. Of these, about 5% failed to respond sufficiently to brief intervention performed with integrity for five to nine days (VanDerHeyden et al., in press).

Common problems associated with screening measures include reliance on a single response opportunity and measurement of student performance without consideration of the “context.” The use of CBM to identify students can be an important first step, but a subsequent time-series analysis allows for comparison of both performance level and slope to that of same-class peers strengthens decision-making. VanDerHeyden and Witt (2003) conducted an examination of the effect of trait and achievement base rates upon the identification accuracy of STEEP and teacher referral. STEEP yielded higher predictive power estimates (i.e., combined positive and negative predictive power) than teacher referral in both high-achieving and low-achieving classrooms. STEEP predictive power estimates in high- and low-achieving classrooms were similar to overall predictive power estimates for STEEP. Relative to STEEP, teacher referral estimates were much more variable, particularly in high-achieving classrooms. Findings included that teachers referred male students disproportionately whereas STEEP did not.

With respect to race, one particularly alarming finding was observed. Greater than 50% of the children who scored in the “at-risk” range on the schoolwide CBM probes (i.e., below the 16th percentiles for their classes and below the instructional range) were African Americans, whereas only 15% of students attending the school were African Americans (VanDerHeyden & Witt, 2003). This finding was particularly alarming since *a greater proportion of African American students responded successfully to intervention*. Hence, the points of comparison indicated that these African American students exhibited a serious academic deficit, but their response to intervention suggested they were not disabled. Further, fall to spring growth on

CBM reading and mathematics measures indicated that children of minority ethnicity (mostly African American) receiving existing instruction made less growth than their Caucasian peers in math and in reading. Yet, when these minority children received an evidence-based intervention their growth was typically comparable to and in some instances greater than their Caucasian peers in that sample.

STEEP and Students who are English Language Learners (ELL)

The task of meaningful and unbiased assessment during the identification process for special education has recently become even more difficult with the growing number of children entering our schools with limited English proficiency. Many standardized tests that assess the ability to learn are technically inadequate or result in test scores that have different meanings when given to minority language students (Shinn, 1998). Concern over the limited validity of assessment practices for ELL students has grown largely due to the problem of overrepresentation of language minorities in special education programs (Artiles & Trent, 2000; National Center for Learning Disabilities Policy Brief: Minority Students in Special Education, 1999). Selection bias becomes an important issue if students are over-referred for placement in special education, especially if placement is more of a risk than a benefit for some children. For example, ELL students participating in special education may be in classrooms with greater heterogeneity of students that may stifle language development and participation in these classes may result in less exposure to general education curriculum (Artiles & Trent, 2000). More importantly, very little is known about ELL students with disabilities and we have little information about effective individual interventions for these children (August & Hakuta, 1997; Gersten & Baker, 2000). In fact, after 3 years of special education, Wilkinson and Ortiz (1986) reported that “LD” Hispanic students failed to make academic gains after a 2-year period and

achievement scores were at the same level as at entry in special education program. Hence, a score used to identify eligibility to special education may have potential negatives consequences due to the environmental context of the special education classroom.

Recently, there is some evidence that under referral of ELL students instead of over referral is becoming a problem in some school districts (Gersten & Woodward, 1994). The tendency toward under referral may be partly due to the common practice within some schools to wait until students have reached a certain level of English proficiency before evaluating if further educational assistance is needed. However, peer comparison of slope estimates that are monitored as students simultaneously are learning academic content and learning English language will provide schools with an estimate of typical academic growth of learning for ELL within their local curriculum.

Although the STEEP screening model has been evaluated with African American populations, such evaluation efforts have not been conducted with ELL students. Two challenging issues that are unique to this population will be confronted when evaluating the effectiveness of the gated STEEP system. First, school assessment teams are often confronted with the task of determining whether a bilingual student's lack of academic progress is due to language ability or due to a disorder that may lead to failure without effective services. For ELL students to be considered *possibly* disabled they would also need to be different in 3 ways: level of performance, slope, and intervention slope. Due to potential language limitations, many ELL students will most likely not match English speaking peers. Given the disproportionate number of minorities in special education, accuracy of eligibility decisions may be enhanced if achievement rates with minority students are also compared to students with similar backgrounds. Sole reliance on the class norm may not fully represent actual abilities since a

small number of ELL students in a classroom may not be large enough to exert any influence on the class mean. Thus, including an evaluation of comparisons of level and slope with peers who have similar language barriers within the local curriculum is critical. This reference group would represent the child's linguistic and cultural community to help distinguish between temporary difficulties ELL students confront when learning academic skills in a non-proficient language from more chronic deficiencies that interfere with learning. This is not a simple task given that ELL students can differ in language, length of residence in US, language proficiency, and prior school experience (Lam, 1992).

Second, there is a substantial delay in learning English. In general, studies have demonstrated that non-native English speaking students are learning English but it can take two to five years for a child to obtain a fluent level of oral English proficiency (Hakuta, Butler, & Witt, 2000). Hence, educators, like the aforementioned school assessment teams, are faced with the problem of trying to determine if the student's lack of academic growth is due to a disorder that, if untreated, would lead to failure or if it is due to language problems arising from a lack of proficiency in the testing language. A substantial amount of research has consistently demonstrated that early intervention for at-risk learners is more effective than later treatment when the problem has intensified with additional complexities (Donovon & Cross, 2002).

However, we know that there are specific teaching strategies in the classroom that effectively increase growth for all children including ELL students. For example, studies have demonstrated that specifying task outcomes and teaching what students must do to accomplish tasks using demonstrations, providing academic engagement and language (oral and written) practice opportunities with immediate feedback, using physical gestures and visual cues, explaining ideas several times using multiple examples, frequently checking for comprehension,

and monitoring students' progress effectively increase academic progress (August & Hakuta, 2000). Specifically, for ELL, teaching strategies that target academic-related verbal interactions increase both academic and English oral proficiency. If these components are in place, then we increase the likelihood that all children will learn. Hence, context matters at both the school level and classroom level regardless of language.

The results of a class wide assessment presented in Figure 3 provide an illustration of the STEEP procedures with ELL children. This figure displays the reading scores from a schoolwide screening that was administered in a first grade classroom in November. To facilitate decision making, the students' scores are displayed from the lowest to highest reading scores. There are three types of standards that are shown on this graph: a national norm, the class median, and the ELL median. The national norm is a benchmark that students are to obtain by spring semester of first grade (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002). The class median was calculated from reading scores of both native English speaking students and ELL students. To obtain the ELL median, the scores of nine ELL students with similar cultural and language backgrounds were used in this case. Specifically, the nine ELL students were born in the United States, learned Spanish as their primary language, speak Spanish at home, scored within a limited English proficiency range on a proficiency test given at the beginning of the school year, and were receiving ESL services for one hour per day.

The Case of Juan

To illustrate idiographic approaches with ELL children, two cases are presented. Juan was the student with the lowest reading score presented in Figure 3. Juan's performance was compared initially to classmates in order to determine the degree to which he was acquiring reading skills in English. When Juan was given the performance/skill deficit evaluation to

determine if incentives helped, there was no increase in his performance. According to the STEEP decision-making model, these results suggest that Juan was exhibiting a skill deficit.

Before concluding that Juan was having severe academic difficulties, his score was compared to other students of similar cultural and language background in his class to examine the influence of learning a second language on reading performance. The data presented in Figure 3 shows that most students are progressing towards the national norm, however, the ELL median is lower than the class median score. Comparing his score to the median score of ELL students shows that Juan's score fell substantially below this standard. With this information alone, the accuracy with which these scores can be interpreted without bias is a bit limited. One interpretation may be that Juan has a disability since he is performing below other native Spanish speaking students. However, Juan just moved to the district in November. Thus, an alternative interpretation of this score may be that Juan did not have adequate instruction, if any, on *this skill* at his other school. Inadequate instruction is a plausible explanation of Juan's performance since Juan may have attended a school with little content overlap with the present school or he may have received inconsistent bilingual/ESL services. Without additional information all that can be said about Juan's score is that it is reflective of his performance at one point in time and suggests only that Juan's current performance is lower than expected if he is to have success in this subject. It should be noted too that it would be inappropriate to conclude from the comparison between Juan's level of performance and his peers that Juan's lower score is a reflection of Juan's inability to learn. The results displayed in Figure 3 do not indicate why he scored lower than his classmates nor can it be said that he would continue to do so.

Some may suggest that, as an alternative assessment strategy, information be obtained about Juan's previous academic experience; however, this can be time consuming and is limited

by the fact that no direct observation of the quality or intensity of his past learning experiences is possible. Since it is difficult to ascertain a student's past learning history, a simple evaluation of Juan's performance when given empirically supported instruction was conducted. Figure 4 shows Juan's performance during intervention. The data presented in this figure show the results obtained when Juan and five ELL students were given a small group reading intervention consisting of a review of key vocabulary followed by listening passage preview and oral practice (Rousseau, Tam, & Ramnarain, 1993). After two weeks of small group instruction, Juan's score improved at a faster rate than his peers. Although it is not clear why Juan failed to perform at his previous school, this information suggests that Juan benefits from the type of instruction that is given in the present classroom. An examination of Juan's intervention progress suggests that his low score in November was perhaps more attributable to lack of effective instruction than a disability. Given that his current progress in the classroom during intervention is good, this helps to rule out the possibility of a severe learning disability and these data can be used to predict that he would not need additional services.

The case of Renee

The results for the schoolwide screening in Figure 3 indicated that Renee, also an ELL child, was also not benefiting from instruction. Unlike Juan, Renee has attended her present school since kindergarten; hence, she has been given the same instruction and curriculum as her classmates. Thus, Renee's reading performance is compared to an ELL group representative of students who are learning English and reading within the same local context. The peer comparison data showed that her level was lower than both the class and ELL median scores.

The question of whether Renee has received adequate instruction should still be considered. By assessing more closely we can examine alternative explanations for her low level

of achievement which center around whether her learning difficulties are due to insufficient exposure to English rather than pervasive learning problems. One potential explanation of her poor performance is that learning experiences within the classroom may have provided her with too few opportunities to practice skills needed for her to achieve. For example, a key component for academic growth for ELL students is frequent verbal and written practice of skills (Gersten & Baker, 2000). However, these opportunities may not be equally distributed across students. Several studies have found that there are very few opportunities for oral responses in most classrooms classroom and that teacher/student interactions which do occur between ELL students and teachers were teacher-initiated or teachers asked questions that required few words or recall (Arreaga-Mayer, & Perdomo-Rivera, 1996; Lopez-Reyna, 1996). Moreover, teachers often decrease language expectations and reduce details when academic demands are high. As a result of this, some students may be presented with an over-simplified curriculum without pertinent background knowledge (August & Hakutu, 1997). Hence, even within the context of the classroom, Renee may be given different types or amounts of response opportunities that may to some degree explain her academic difficulties.

One method for assessing this possibility is to directly measure and examine Renee's reading progress when given the same number of learning trials and response opportunities as her peers during an intervention. Results of this type of intervention, following baseline on a reading skill are shown in Figure 4. The intervention did not result in substantial improvements in level or slope. Thus, this process provided data that showed that Renee did not benefit from classroom instruction or small group instruction in comparison to her English speaking and ELL classmates. For Renee, there may be some other instructional factors that are related to second-language issues that would explain the low score. Her rate of English adoption may have

differed from her peers due to individual differences in language dominance or background knowledge. Thus, Renee may not have the language tools necessary to learn to read at the same rate as other ELL students. Although language cannot be ruled out as a primary explanation for Renee's reading difficulties at this time, this process distinguished that Renee is in need of an intense change in instruction when compared to other students who are learning to read in English.

These two cases illustrated an evaluation of academic progress for students who were not yet proficient in the language of testing. These types of measures can play a powerful role in the way educational decisions are made for language minority students. Instead of waiting 2 to 5 years for English proficiency, a peer comparison at a single point in time and an examination of progress over time provided an immediate picture of student progress on critical academic skills. Although the ELL students' skill levels were lower than English speaking peers, a comparison of scores with a reference group that represents the child's linguistic and cultural community provided some evidence that further helps us to distinguish the temporary difficulties ELL students face in learning to function in a non-proficient language from more permanent deficiencies that interfere with learning.

Summary

Race, ethnicity, and culture are integral parts of the human experience that influence how we view the world, ourselves, and others. They influence what we know, do, and believe. These group difference variables (i.e., culture) exist in a complex overlay with the diversity of individual differences that influence both academic accomplishment and social functioning. Although the relative importance of experiential or genetic differences in creating these individual differences is frequently unclear for individuals, their synergistic influence is evident.

As schools have attempted to make sense of and respond to the tremendous individual diversity they have relied on traditional conceptions of assessment within psychology and education.

These conceptions are nomothetic.

Nomothetic assessment's emphasis on norms and the discrimination of reliably different groups has added to the societal interplay between individual differences and culture. Suspicion and allegation have emerged that schools' nomothetic assessment practices are biased and unfair. These concerns have focused on issues such as disproportionate representation in segregated special programs and in limited access to additional educational supports. It appears that the traditional nomothetic emphasis on establishing new groups and in this case frequently disability groups may have further exasperated the societal tensions between concerns over equity across racial, ethnic, and cultural groups and services for students who are not succeeding at school.

This chapter described an emerging general approach to this problem and a specific implementation of this approach that are described as eco-behavioral. This approach is based on an extensive and venerable literature that has emphasized the incorporation of idiographic methods of assessment and decision making for individual students. The approach emphasizes a time efficient decision-making model that is flexible, treatment oriented, repeatable, and places minimal emphasis on the establishment of new groups of students. The approach emphasizes a three-anchor comparison system for evaluating individual differences. These points of comparison are local norms, criterion referenced standards, and response to intervention.

The general approach described in this chapter is being implemented in varied permutations in sites across the United States. Although the evaluation literature in this area is incomplete, there appears to be increasing interest in contextualized assessment and resistance to intervention as anchors for the identification of and provision of services for students who

exhibit specialized need. Although the ultimate success of this approach in coping with both society's concerns for equity and students' unique needs is as yet unknown, one thing appears nearly certain. Continuing to approach diversity of culture and individual needs in the same manner is likely to perpetuate the same conflicts and concerns that have been evident for some time.

References

- Afridi, S. (2001). *Muslims in America: Identity, diversity, and the challenge of understanding*. New York: Carnegie Corp. of New York.
- Arreaga-Mayer, C. & Perdomo-Rivera, C. (1996). Ecobehavioral analysis of instruction for at-risk language–minority students. *Elementary School Journal*, *96*, 245-258.
- Artiles, A. J., & Trent, S. C. (2000). Representation of culturally/linguistically diverse students. In C. R. Reynolds, & E. Fletcher-Jantzen (Eds.), *Encyclopedia of special education* (pp. 513-517). New York: John Wiley & Sons.
- Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education*, *27*, 410-437.
- August, D. & Hakuta, K. (1997). *Improving schooling for language minority students: A research agenda*. Washington, DC: National Academy Press.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*, 91-97.
- Bahr, M. W., Fuchs, D., Stecker, P. M., & Fuchs, L. S. (1991). Are teachers' perceptions of difficult-to-teach students racially biased? *School Psychology Quarterly*, *20*, 599-608.
- Barbetta, P. M., Heward, W. L., Bradley, D. M., & Miller, A. D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of sight words of students with developmental disabilities. *Journal of Applied Behavior Analysis*, *27*, 177-178.
- Barnett, D. W., Bauer, A. M., Ehrhardt, K. E., Lintz, F. E., & Stollar, S. A. (1996). Keystone targets for change: Planning for widespread positive consequences. *School Psychology Quarterly*, *11*, 95-117.

- Bau, A. M. (1999). Providing culturally competent services to visually impaired persons. *Journal of Visual Impairment & Blindness, 93*, 291-297.
- Bell, P. F., Lentz, F. E. & Graden, J. L. (1992). Effects of curriculum-test overlap on standardized test scores: Identifying systematic confounds in educational decision making. *School Psychology Review, 21*, 644-655.
- Bennett, K., & Cavanaugh, R. A. (1998). Effects of immediate self-correction, delayed self-correction, and no correction on the acquisition and maintenance of multiplication facts by a fourth-grade student with learning disabilities. *Journal of Applied Behavior Analysis, 31*, 303-306.
- Cole, N. S. (1981). Bias in testing. *American Psychologist, 36*, 1067-1077.
- Daly, E. J., Lentz, F. E., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11*, 369-386.
- Daly, E. J., & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27*, 459-469.
- Daly, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554-574.
- Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.
- Donovan, S.M. & Cross, C.T. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- Elliott, S. N. (1991). Authentic assessment: An introduction to a neobehavioral approach to

- classroom assessment. *School Psychology Quarterly*, 6, 273-278.
- Fletcher, J. M., Francis, D. J., Shaywitz, S. E., Lyon, G. R., Foorman, B. R., Stuebing, K. K., & Shaywitz, B. A. (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice*, 13, 186-203.
- Fletcher, J. M., Lyon, G. R., Barnes, M., Steubing, K., Francis, D., Olson, R., Shaywitz, S., & Shatwitz, B. (2002). Classification of learning disabilities: An evidence-based evaluation. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research into practice* (pp. 185-250). Mahwah, NJ: Lawrence Erlbaum, Inc.
- Fuchs, L. S., & Deno, S. L. (1982). *Developing goals and objectives for educational programs*. Washington: American Association of Colleges for Teacher Education.
- Fuchs, L. S. & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, 13, 204-219.
- Gersten, R. & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children*, 66, 454-470.
- Gersten R. & Woodward, J. (1994). The language-minority student and special education: Issues, trends, and paradoxes. *Exceptional Children*, 60, 310-322.
- Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly*, 11, 326-336.
- Good, R. H., Simmons, D. S., Kame'enui, E. J., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University

of Oregon.

- Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly, 12*, 249-267.
- Hakuta K., Butler, Y. G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* The University of California Linguistic Minority Research Institute Policy Report 2000-1. Retrieved April, 2000 from <http://www.stanford.edu/~hakuta/Docs/HowLong.pdf>
- Haynes, S. N., & O'Brien, W. H. (1999). *Principles and practice of behavioral assessment*. New York: Kluwer.
- Hayes, S. C. Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*, 963-974.
- Howell, K. W., Fox, S. L., & Morehead, M. K. (1993). *Curriculum-based evaluation: Teaching and decision making*. Pacific Grove, CA: Brooks/Cole Publishing Co.
- Ikeda, M. J., Tilly, W. D., III, Stumme, J., Volmer, L., & Allison, R. (1996). Agency-wide implementation of problem solving consultation: Foundations, current implementation, and future directions. *School Psychology Quarterly, 11*, 228-243.
- Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984-1024). New York: Wiley.
- Kelley, M. L., Reitman, D., & Noell, G. H. (2002). *Practitioner's guide to empirically based measures of school behavior*. New York: Kluwer.

- Lopez-Reyna, N. (1996). The importance of meaningful contexts in bilingual special education: Moving to whole language. *Learning Disabilities Research & Practice, 11*, 120-131.
- Lam, T. C. (1992). Review of practices and problems in the evaluation of bilingual education. *Review of Educational Research, 62*, 181-203.
- Macmann, G. M., & Barnett, D. W. (1999). Diagnostic decision making in school psychology: Understanding and coping with uncertainty. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 519-548). New York: Wiley.
- MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998). Discrepancy between definitions of learning disabilities and school practices: An empirical investigation. *Journal of Learning Disabilities, 31*, 314-326.
- Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In S. N. Elliot & J. C. Witt (Series Eds.) & M. R. Shinn (Volume Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 18-78). New York, NY: Guilford Press.
- McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc cognitive abilities research on the future use and interpretation of intelligence test batteries in the schools. *School Psychology Review, 26*, 189-210.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*, 741-749.
- Noell, G. H., Duhon, G. J., Gatti, S. L., & Connell, J. E. (2002). Consultation, follow-up, and behavior management intervention implementation in general education. *School Psychology Review, 31*, 217-234.

- Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39*, 335-355.
- Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*, 77-88.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed). New York: McGraw-Hill.
- Quiroga, T, Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Beringer, V. W. (2002). Phonological awareness and beginning reading in Spanish-speaking ESL first graders: research into practice, *Journal of School Psychology, 20*, 85-111.
- Reschly, D. J., & Bersoff, D. N. (1999). Law and school psychology. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 1077-1112). New York: Wiley.
- Reschly, D. J., Kicklighter, R., & McKee, P. (1988). Recent placement litigation: II. Minority EMR overrepresentation: Comparison of Larry P. (1979, 1984, 1986) with Marshall (1984, 1985) and S-1 (1986). *School Psychology Review, 17*, 22-38.
- Rhymer, K. N., Skinner, C. H., Henington, C., D'Reaux, R. A., & Sims, S. (1998). Effects of explicit timing on mathematics problem completion rates in African-American third-grade elementary students. *Journal of Applied Behavior Analysis, 31*, 673-677.
- Roe, M. F., & Vukelich, C. (1998). That was then and this is now: A longitudinal study of teachers' portfolio practices. *Journal of Research in Childhood Education, 12*, 16-26.
- Rourke, B. P., & Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning: Perspectives from neurology and neuropsychology. *Journal of Learning Disabilities, 30*,

34-46.

- [Rousseau, M. K.](#), Tam, B., & [Ramnarain, R.](#) (1993). Increasing reading proficiency of language-minority students with speech and language impairments. *Education & Treatment of Children*, 16, 254-271.
- Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.
- Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York, NY: Guilford.
- Shinn, M. R. & Bamonto, S. (1998). Advanced applications of Curriculum-Based Measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York, NY: Guilford Press.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, C. H., Cooper, L., & Cole, C. L. (1997). The effects of oral presentation previewing rates on reading performance. *Journal of Applied Behavior Analysis*, 30, 331-333.
- Snow, C. E., Burns, M. S., & Griffin, P. (1999). *Preventing reading difficulties in young children*. National Research Council, Committee on Prevention of Reading Difficulties in Young Children. Washington, DC: National Academy Press.
- Sonnenshein, S., & Munsterman, K. (2002). The influence of home-based literacy reading interactions on 5-year-olds' reading motivations and early literacy development. *Early Childhood Research Quarterly*, 17, 318-337.
- VanDerHeyden, A. M., & Witt, J. C. (2003). Quantifying the context in assessment: Capturing the effect of base rates on screening accuracy. Manuscript submitted for publication.
- VanDerHeyden, A. M., Witt, J. C., Naquin, G. (in press). Development and validation of a process for screening referrals for special education. *School Psychology Review*.
- Vellutino, F. R., Scanlon, D. M., & Lyon G. R. (2000). Differentiating between difficult-to-

remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33*, 223-238.

Wilkinson, C., & Ortiz, A. (1986). *Characteristics of limited English proficient and English proficient learning disabled Hispanic students at initial assessment and reevaluation*. Austin University of Texas, Department of Special Education, Handicapped Minority Research Institute. (ERIC Document Reproduction Service No. 283 314).

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher usage of interventions in general education: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693-696.

Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Principles and procedures of applied behavior analysis with exceptional children*. Boston: Allyn & Bacon.

Wolf, 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.

Figure Captions

Figure 1. This figure shows the performance of each child in a given class during the schoolwide screening. Each bar represents the number of digits completed correctly in two minutes for all children in the class on the math task. The lowest-performing children participated in a Performance/Skill Deficit Assessment. The second bar for these children represents the score obtained given the opportunity to earn a powerful incentive. Three shaded areas in the background (provided in color in schools) indicate whether performance falls in the frustrational range (lowest shaded area), instructional range (middle shaded area), or mastery (highest shaded area).

Figure 2. Some of the variables impacting intervention effectiveness.

Figure 3. Data shown are from a case example of a schoolwide screening in reading with a group of students who are English Language Learners. Each gray bar represents the number of words read correctly in one minute. The lowest-performing children participated in a Performance/Skill Deficit Assessment. The second bar for these children represents the score obtained when given the opportunity to earn a reward. The dotted lines indicate performance at or above a national norm, the median score of all students, or the median score with students who are ELL.

Figure 4. Growth trends for individual children who are English Language Learners with a skill deficit in reading during reading intervention combining a review of Key words with a listening passage preview.

Figure 1.

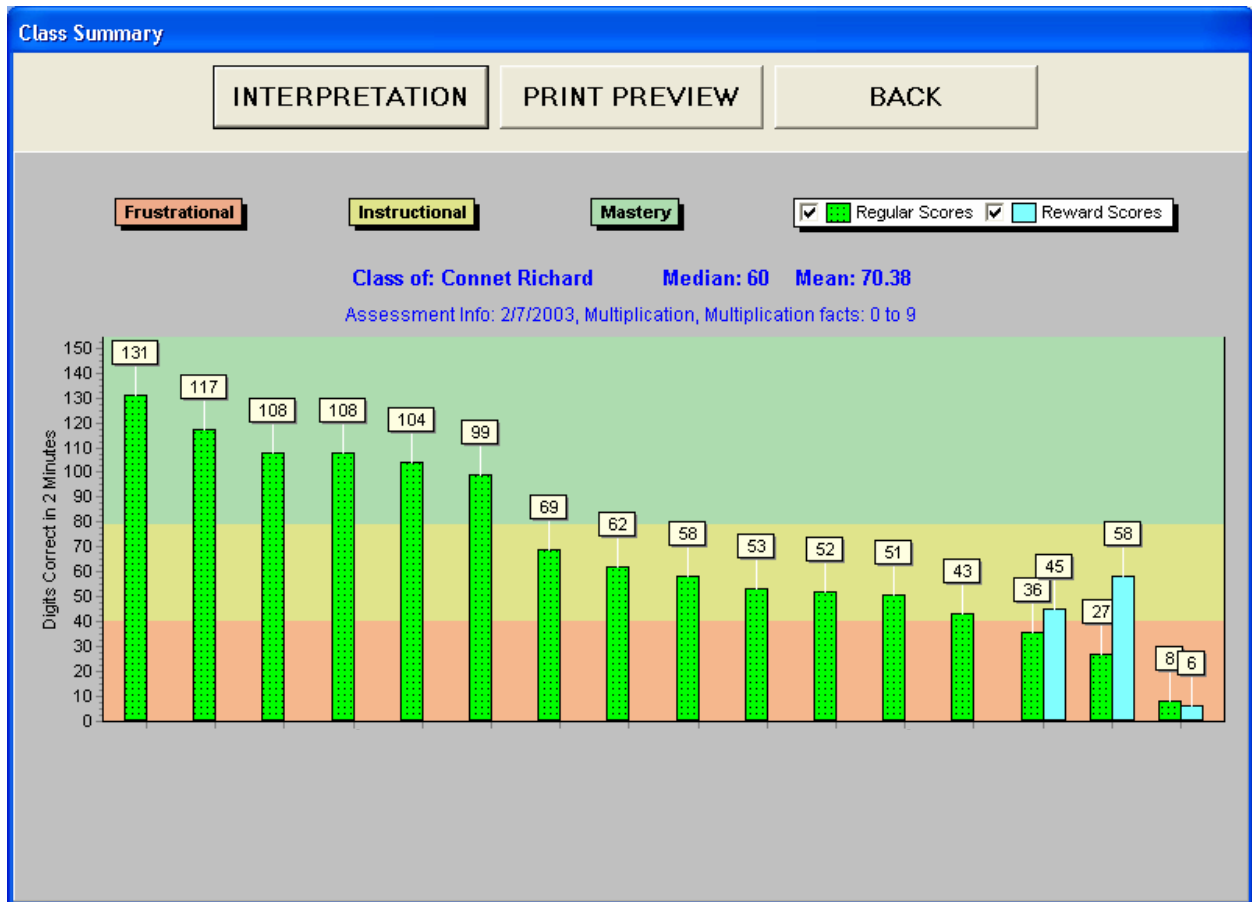


Figure 2.

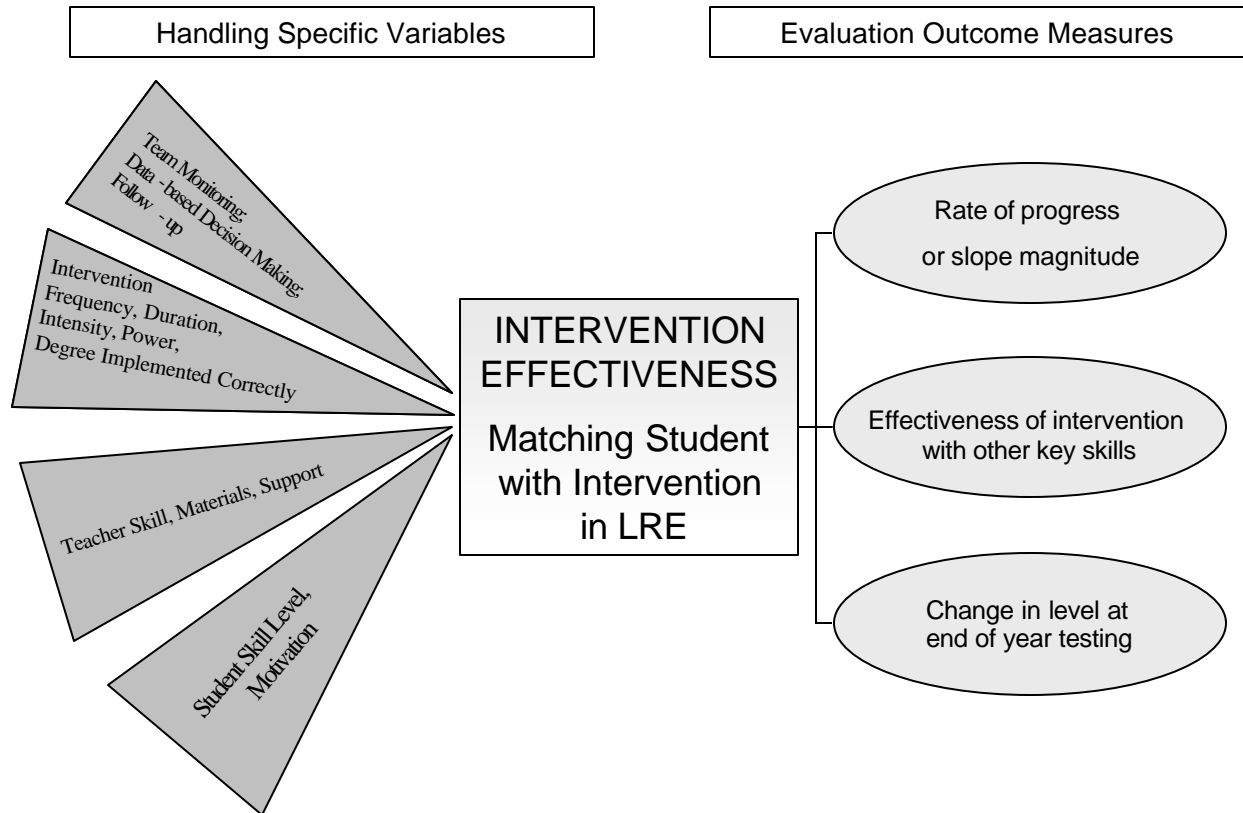


Figure 3.

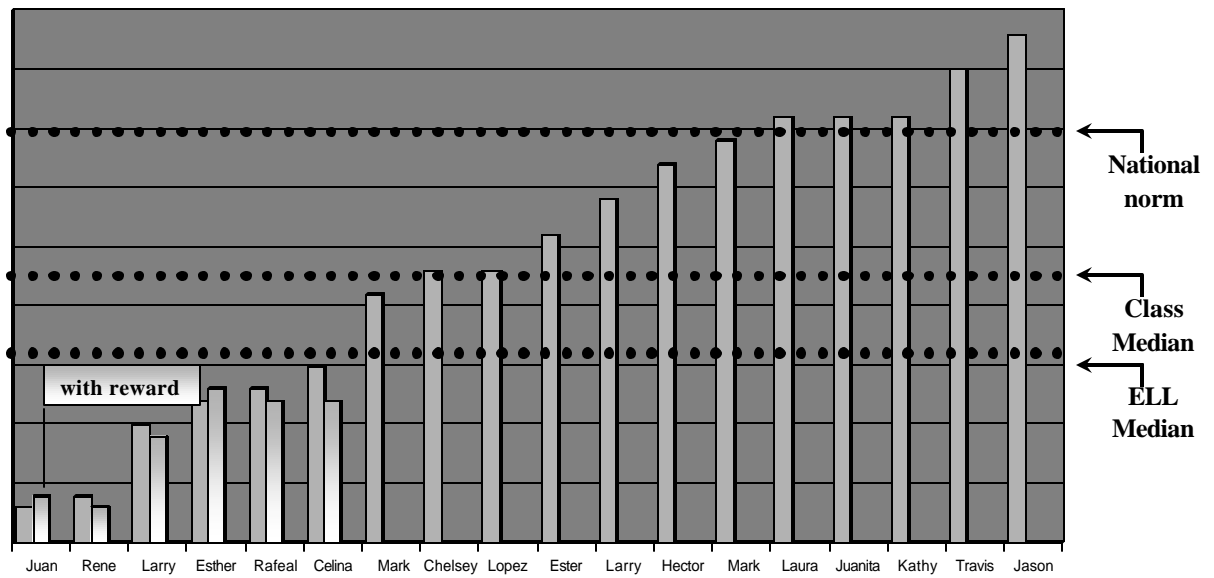


Figure 4.

