Running head: CHARTING FUTURES

5/11/04

Achieving Science-based Practice through Response to Intervention: What it Might Look Like in Preschools

David W. Barnett

University of Cincinnati

Amanda M. VanDerHeyden

Vail School District

Joseph C. Witt

Louisiana State University

Notes: Appreciated is extended to Janet Graden and Ed Lentz who helped develop the definitions and permanent products for a related RTI model for school-age children, and to Maria Oddo who helped develop the graph. Please correspond with David Barnett, PO 210002, University of Cincinnati, Cincinnati, OH 45221-0002, David.Barnett@UC.EDU.

Abstract

Decades-old controversies still surround service delivery decisions and intervention qualities supporting children's progress within typical settings. Although the performance of all children within an agency is of interest, at the center of controversy is the group of children who may require specialized or intensified instruction, goals, or supports, and the associated professional practices used for decision making. The purpose of this review is to examine *response to intervention* (RTI) as applied to preschool services: class-wide (Tier 1), embedded or small group (Tier 2), and intensive individualized interventions that may include disability services (Tier 3). First, two conflicting directions in measurement underlying special services decisions are reviewed: 'traditional' and 'direct' or 'intervention-based' assessment. Second, a synthesis and recommendations are provided based on a unified response to intervention model for preschool service delivery. Third, to address accountability, *permanent products* of intervention-based services are used to map a child's path through tiers of services: Last, challenges to RTI are discussed.

Achieving Science-based Practice through Response to Intervention: What it Might Look Like in Preschools

Professionals and researchers in early childhood have discussed, sometimes with considerable tension, the need for practices with documented effectiveness in meeting the needs of children with concerning performances and challenging behaviors (Carta, Schwartz, Atwater, & McConnell, 1991). The debates are far from settled (Kohn, 2001; Strain & Joseph, 2004), and, in fact, instructional questions may have become even more complicated with recent mandates pertaining to achievement such as No Child Left Behind Legislation (PL 107-110) and Head Start Child Outcomes Framework (U. S. Department of Health and Human Services, 2000). In the past, a relatively small group of children with disability questions about behavior or predicted academic performance were the major focus in considering decision making and programming for special needs. At present, in Head Start as well as in public education, goal setting, progress monitoring, and instructional quality are targets of scrutiny for all children within a school or agency. Response to intervention (RTI) has evolved from applied behavior analysis and disability research with the potential for serving as a *general* model organizing these service delivery questions for Head Starts or preschools that include typical children as well as those at risk or identified as having a disability.

The purpose of this review is to consider the possibility of the RTI model for preschool services and to suggest professional and agency practices that fit a strong model. To accomplish this objective, we define RTI and clarify decision making differences between traditional disability evaluation and response to intervention. Following a review of RTI models,

recommendations are made for integrating typical preschool targets associated with reform agendas, problem solving, and disability evaluation and services, into a tiered model of preschool service delivery. Last, we discuss challenges to a full RTI model.

What is Response to Intervention?

Response to intervention uses the quality of student responses to empirically-supported interventions as the basis for decisions about needed services. A response to intervention model has a simple structure from applied behavior analysis (ABA). Targets and goals for instruction or intervention are described, such as accelerated academic growth or increased positive social interactions. Interventions with empirical support are introduced and are monitored with frequent measurement of children's performance as well as adherence to the intervention. Children who do not adequately respond to the intervention receive a more intensified or stronger intervention, with measurement and problem solving continuing as needed until an intervention is found that is effective. Following success, goals include reducing intervention intensity as much as possible consistent with the ultimate objective of natural educational programming. What RTI adds to the structure from ABA is that data from *estimates of needed intervention quality and intensity can be used to help answer service delivery questions* as opposed to the use of estimates of child deficits from test-based results.

A full response to intervention model appropriate to preschools and Head Start services, and elaborated in subsequent sections, can be organized by *tiers* of services beginning with school- and class-wide (Tier 1), 'embedded' or small group interventions for 'non-responding' children (Tier 2, traditionally referred to as remedial), and intensive and individualized problem solving (Tier 3). Ideally, tiers are *added* as necessary. Screening and problem solving begins in

Tier 1, and if that proves to be insufficient, Tier 2 interventions are integrated into a child's day. If goals are not met with Tiers 1 and 2 services, Tier 3 planning occurs which may include all tiers simultaneously. For each tier, teams systematically evaluate interventions that may progressively increase (or decrease) in intensity, and they make data-based service delivery decisions from results (e.g., Tilly, Reschly, & Grimes, 1999).

Thus, response to intervention is characterized by empirically-based, systemic, and structured problem-solving procedures that may be used for making systems-level as well as individualized child-level educational decisions. The beginning and ending criterion is success in typical environments. The core features of single case designs (baseline, intervention; Wolery & Dunlap, 2001**)** are used for evaluating interventions across a continuum of intervention intensity that underlies response to intervention.

Why Response to Intervention?

Response to intervention is a set of empirical tactics that directly address questions of *optimal intervention intensity and design* for children. Despite the many challenges, it is what needs to happen for many children. RTI uses step-by-step or *graduated* increases or decreases in intensity to demonstrate the initial, ongoing, and ultimately least restrictive intervention qualities for a child. Intervention intensity can be defined in various ways but fundamentally it reflects the time, effort, or resources that, as intensity increases, are difficult to support without specialized services. Response to intervention yields a *data set* based on careful analysis of student contextual performance and needs. Children progressing through systematic iterations of response to intervention and requiring considerable intensity and support for success may be

candidates for services provided through special education.

Legislation and Major Initiatives Supporting Response to Intervention

First, IDEA '97/'99 (PL105-17) provides a legal basis and specific guidelines for what may be a consensus in effective innovation and reform (Reschly & Ysseldyke, 2002): an intervention-based, needs-driven, team approach to serving children with disabilities that emphasizes prevention efforts. At the heart of the *findings* (sec. 601. c, pp. 38-42) are key themes that set the stage for fundamental reform efforts: (a) improving methods of setting goals; (b) emphasizing "proven methods of teaching and learning"; (c) encouraging "whole-school approaches and pre-referral interventions" to reduce errors in identification; (d) conducting step-by-step contextual and functional assessments; and (e) applying the construct of positive behavioral support for children with disabilities whose behaviors seriously interfere with learning (Carr, Horner, & Turnbull, 1999; Turnbull, Wilcox, Stowe, Raper, & Hedges, 2000). Response to intervention concepts are found in *No Child Left Behind* legislation (PL 107-110) that, among many science-based program components, requires effective instruction and progress monitoring as a starting point for educational accountability before special education placement. Recent legislation supports a sequential problem-solving process founded in science for making educational decisions. Eligibility determination based on RTI is evidenced by technically adequate progress monitoring data within a set of stringent decision criteria. Evaluating data in terms of its step-by-step or *incremental validity* to special services eligibility is consistent with the language and intent of IDEA (SEC. 614 [c]): "As a part of an initial evaluation...the IEP Team...shall (A) review existing evaluation data on the child, including evaluations and information provided by the parents of the child, current classroom-based

assessments and observations, and teacher and related services providers observation; and (B) on the basis of the review, and input from the child's parents, identify what additional data, if any, are needed to determine..."[disability-related questions] (1997, 111 stat. 39).

The report of the President's Commission on Excellence in Special Education (OSERS, 2002) provides extensive support for the above conclusions and makes important recommendations for change. The report recommends the abandonment of the traditional classification process in favor of decisions based on response to instruction for specific learning disabilities (SLD). A growing body of literature suggests the general promise of this idea for disability decision making (e.g., Barnett, Bell, et al., 1999; Barnett, Daly, Jones, & Lentz, in press; Gresham, 1991).

Second, scientifically validated, continuous progress monitoring (Fuchs & Fuchs, 1986; Fuchs & Fuchs, 1998) is strongly emphasized for making instructional decisions that lead to effective special services. Third, service delivery models should not be based on 'waiting for failure' before organized interventions are attempted (e.g., O'Shaughnessy, Lane, Gresham, & Beebe-Frankenberger, 2003; Turnbull et al., 2002). Fourth, given sufficiently documented and technically-adequate data, RTI as a child's *cumulative intervention history* may be used to help determine the need for specialized services as required by IDEA (PL 105-17, sec. 614[c]).

Different Paths for Educational Decisions

Whereas conflicting arguments for educational decisions may be founded in philosophy or developmental theory (e.g., Carta, et al., 1991), in practice, one of the greatest obstacles in actualizing the intent of 'response to intervention reform' is needed change in the measurement

practices associated with children's educational performances and challenging behaviors. Much of what currently occurs in measurement has no link to developing or monitoring effective educational practices and the above goals of reform (Neisworth & Bagnato, 1992). This section represents, for the unconvinced, 'why a need for change?' Whereas some of the criticism is well-traveled, many of the critics have not presented a full set of concerns.

The 'psychometry' of practice and 'cut score magic.' A major underpinning of reform effort has been the analysis of data used for decision purposes. Data use has typically centered on children's classifications to ostensibly determine needs, services, and outcomes. There is no 'good news' about individualized test use when scrutinized by the lens provided by decision analyses of typical children's one-point-in-time test or profile scores. Despite many emphatic calls for change (e.g., Barnett, Macmann, & Carey, 1992; Galagan, 1985; Neisworth & Bagnato, 1992; Reschly, 1988), many new instruments (or new interpretations of old instruments) have been introduced. Perhaps the most damaging outcome of instrument development and promotion within the context of reform has been the diversion of professional attention and resources away from effective contextual measurement and intervention activities (Reschly & Ysseldyke, 2002).

The key arguments by proponents of new measurement tools have been that new theories and measures are needed and might be useful, or that new methods of analysis may ameliorate concerns with score interpretation (e.g., Keith, Krantzler, & Flanagan, 2001; Kush et al., 2001). No reasonable professional would argue with these assertions about the evolution of constructs and measures that may ultimately benefit children. However, test and scale development will not progress *without* addressing two significant criticisms pertaining to decision analyses. These criticisms also have a long history, but seem to be an unfortunate backdrop to test development

and use rather than a large part of the ultimate criteria for judging the soundness of instrument use. Some test critics have developed one criticism or the other, but the questions they represent work together in empirical and professional concert.

First, among the most well documented concerns are what seem to be *intractable* problems associated with achieving professionally defensible levels of *decision reliability* for common one-point-in-time measures of constructs (i.e., cognitive skills) used for *individual* educational diagnostic decisions and confidence in results (Barnett, Lentz, & Macmann, 2000). Second, *decision validity* raises the question of whether or not some benefit accrues for children that directly stems from test-based inferences (i.e., that outcomes will be helpful). The lack of demonstrated utility for decision-making and intervention with one-point-in-time assessments continues to be a pertinent question (e.g., Fletcher et al., 2001) that has not been addressed by test proponents (Hayes, Nelson, & Jarrett, 1987). Perhaps unfortunately, the ideal of *consequential validity* (Messick, 1988) and the term *treatment validity* itself have connoted some misunderstandings about what desirable criteria might be (e.g., Alfonso, Oakland, LaRocca, & Spanakos, 2000). Major reviews citing the benefits of scales do so without *any* evidence of benefit, but rather imply some benefit often based on only modest correlations within or between similar measures (i.e., other developmental or achievement scales) (Meyer et al., 2001). The idea of examining test use consequence has not yet evolved to be a burden on test developers, but the requirement of addressing consequences *is* a part of professional practice (AERA, APA, & NCME, 1999). A lack of attention to misguided consequences of test use may leave a professional, parent, and child with the expensive burden that test developers and publishers

successfully avoided.

We use commercially-developed tests of intelligence to represent one-point-in-time measures because their scores are widely regarded by many as having adequate technical characteristics and as a fundamental achievement of social science. Here are recent examples of *consequence* assertions with which children and professionals would have to live. “Valuable information from well-constructed measures,...as the WISC-III, *could* be used to predict the likelihood of intervention success” (Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001, pp. 50-51, emphasis added); “we...*believe* that measures of intelligence...have much to offer...and can ultimately be of great benefit to the children who complete them” (Kush et al., 2001, p. 85, emphasis added). These are expensive assertions for social scientists to make without any supporting data, especially after 100 years of research efforts (Wolf, 1973). The instability of *decisions* based on IQ estimates for young children, along with the fallacy of topography or category-based treatments, particularly when the sorting of children into groups is not a reliable process, make test-based services highly questionable. Further, findings indicate that IQ is not a useful predictor of response to intervention (Gersten, Becker, Heiry, & White, 1984; Vellutino, Scanlon, & Lyon, 2000; Vellutino, et al., 1996; Vellutino, Scanlon, & Tanzman, 1998).

The measurement questions that we raise can apply to *any* one-point-in-time measure and decision context because they stem from the theoretical behavior of correlated variables and proper error estimates, not from the type of scale (i.e., behavior rating, cognitive functioning) (Macmann & Barnett, 1994). Quite logical, devastating for interpretive confidence, and *not* mentioned in test manuals or most reviews, is the fact that traditional estimates of error from reliability facets are *additive* for inferences or decision-making purposes. Error comes from

different sources (e.g., over item content, time, scorers). For proper individualized score interpretation for an instrument-based inference or diagnostic decision, reliability estimates would be used to show bands of error combining *all* possible and relevant reliability coefficients (test/retest, item consistency, scorer, and other) for the decision purpose; all error sources are 'active' for any decision made from a test score. For example, and often resembling a best case scenario, if a diagnostic instrument has an internal consistency reliability of .84 (item content) and a test-retest reliability of .85, the *combined effects* reliability for a decision or inference would be .73, connoting a less than acceptable range of inferential error for high stakes professional decisions (from Leonard Feldt reported in Macmann & Barnett, 1999, p. 531; see also Anastasi & Urbina, 1997, pp. 101-102). Test interpretations for preschool children in most real situations are much more risky than this in terms of error sources, as there is substantially greater error associated with the reliability of profiles or other discrepancy-based interpretations that are not accounted for in traditional presentations of technical adequacy because skills are more variable and less well differentiated for this age group (Barnett et al., 2000).

Thus, it is not the cognitive (or other) theory applied that necessarily constrains effective decision making, a quest that has occupied the past century, but the cumulative nature of errors of inference in common educational or behavioral one-point-in-time measures for individual decisions and omission of context that limits confidence in results. These decision errors are well documented in research but are not evident in cautions for practice. In brief review, decision errors are affected by (a) extreme scores (as scores become more extreme, the absolute amount of measurement score error increases but extreme scores are those most readily interpreted), (b)

multiple unplanned comparisons of correlated measures (difference scores between correlated cognitive and other developmental measures, or subtests on any omnibus scale, are notoriously unreliable and finding a significant difference is partly a function of the number of comparisons made), (c) modest validity evidence inferences from fuzzy or weak constructs (a difference score between two 'known' correlated scales is a newly 'invented' and idiosyncratic construct having a lower reliability than either measure from which it was derived) (e.g., Fiske, 1982), and (d) lack of contextual information for decisions. Routinely, ambiguous difference- or discrepancy-based constructs without *evidence* of usefulness are applied to young individual children *at near chance or low levels* of decision confidence to guide disability evaluation and intervention planning (Macmann & Barnett, 1997). By lack of attention to significant measurement questions, the positive contributions of professionals to benefit children are lost.

Score use for disability-related decisions is guided by various algorithms or *cut scores* to indicate, at one-point-in-time, whether or not a diagnostic label may be applied and that a child qualifies for services. Although a fundamental pursuit underlying disability and service delivery questions, for all of the above reasons, cut score use may be viewed as an inherently precarious professional function for young children for common classifications. There are different scale construction methods, multiple norm groups, often a sea of numbers, but how a scale works in practice, or the act of professional decision making, is not typically reported or researched at all.

What *is* cut score "magic"? Now you see it, now you don't! By changing the school demographics, selection process, or the instrument used, decision analyses and outcomes will likely change as well (Meehl & Rosen, 1955). If different, ostensibly defendable scales measuring the same construct were used on a different day, or by a different examiner, would the

decision be the same? Research suggests that diagnostic profiles may appear and disappear, depending on the time of measurement, rater, scale, risk or delay definition, or other differences in measurement (Macmann, Barnett, Lombard, Belton-Kocher, & Sharpe, 1989; Ysseldyke et al., 1983). Considering that many psychometric decisions are made prior to testing and interpretation, such as which tests and test dates, from the child's viewpoint, critical outcomes such as eligibility and placement may be the result of a chance encounter rather than diagnostic science.

Research (e.g., Gresham, MacMillan, & Bocian, 1998) also indicates that assessment teams often disregard test scores when making classifications. Professional judgment is not a panacea for overcoming the vagaries of test data (Barnett, 1988). Indeed, where professional judgment is substituted for test data, decision errors are also likely because the process has not been operationalized and is very likely to be unstable across settings, teams, and children.

Applying 'confidence' and 'helpfulness' criteria through response to intervention practices. A second possibility exists for identifying needed supports for children and teachers that meets the intent and requirements of *IDEA* and *No Child Left Behind* legislation, and Head Start mandates. Functional and direct assessments of contextual performances and behaviors in response to interventions are used as the basic data set. Even though single data points and decisions are fallible, sharing some of the criticism to traditional scores above, what is significantly different is the contextual pattern over time that emerges for appropriately selected and sensitive measures, and carefully sequenced interventions that are ordered by intensity and carried out correctly. Different technical adequacy criteria apply, as well as many challenges,

reviewed in later sections.

In summary, problems with traditional one-point-in-time diagnostic measures have been ignored by test developers and underestimated in practice. Well-established methods of direct assessment, intervention design, and evaluation, within reach of professionals, may be used to develop a data set to help make service delivery decisions. The data are organized by RTI models.

Response to Intervention Models

A Brief History and Overview of Models for Academic Responding

An early model defined critical elements of response-to-intervention services: (a) criteria for critical deficits in basic skills using repeated measures (CBM probes, direct observations) organized by time series (single subject) design facets; (b) goals that would represent significant progress towards typical classroom expectancies; (c) more intensive services added based on functional need; and (d) special service decisions based on a child's lack of positive response to structured pre-classification efforts (Deno & Gross, 1973). Notable are the use of *functional discrepancies* as the unit of analysis rather than labels, and the utility of the procedures as a general model rather than a specific one (i.e., for Learning Disabilities). These early strategic elements provide a common thread across more recent models that have been appearing in the literature (e.g., Fuchs & Fuchs, 1998; Good, Kame'enui, Simmons, & Chard, n.d.; Vaughn, Linan-Thompson, & Hickman, 2003; VanDerHeyden, Witt, & Naquin, 2003). We elaborate three models that serve as conceptual foundations for a general integrated preschool model.

Fuchs and Fuchs (1998; Fuchs, et al., 2002) described an assessment design for identifying Learning Disabilities (LD). Children considered as LD are *"nonresponders"* to

educational interventions found in typical classrooms identified through several phases (Fuchs et al., 2002, p. 38). The first phase assessments inform decisions about learning rate within the overall instructional environment of classrooms. Given that the instructional setting is adequate, a second phase assessment is used to identify children with dual discrepancies: where level and trend (or rate of learning) are below that of peers. A third phase assessment has the objective of evaluating the degree to which the regular education setting can be adapted to meet individual needs. Special services are considered for a child when general education interventions do not lead to significant gains. In a fourth phase, a "diagnostic special education trial" is used to empirically determine the contributions of special education for the individual child (p. 43). This may also lead to extended diagnostic trials and more intensive or specialized services.

A model developed by Witt and colleagues (VanDerHeyden, et al., 2003; Witt, Daly, & Noell, 2000) described as *Problem Validation Screening* (PVS) has the following components: "brief classwide curriculum-based measurement (CBM) screenings, direct observation, comparison to same-class peers (as well as national standards, *added),* assessment of the effect of providing an incentive for improved performance (to determine a skill or performance deficit, *added),* and intervention conducted in the natural setting and monitored for procedural integrity" (VanDerHeyden, et al., 2003, p. 206). A referral for a full evaluation is based on lack of positive response to a correctly carried out intervention in the regular classroom.

Positive Behavioral Supports: A Model for Social/Life-Style Outcomes

Academic RTI models have many similarities to the most prominent educational model for social behaviors, *positive behavior support,* which also has origins within applied behavior

analysis and developmental disability research, with objectives directed to *quality of life*. Positive behavior support is described by a continuum of 3 tiers consisting of universal support and primary prevention, specialized group support and secondary prevention, and specialized individual support or tertiary prevention (Sugai, et al., 2000; Turnbull, et al., 2002). Universal interventions are systemic, both school- and class-wide, and for all children, and are relatively low on the continuum of intervention intensity. Specialized group and individual supports increase in intensity and target decreasing numbers of children.

An Integrated Model for Preschool Response to Intervention

There are specific reasons for a unified preschool RTI model for preschool services: (a) testing controversies are not specific to a diagnostic category (i.e., LD) and similar controversial issues can be raised within any behavioral or developmental category or classification outlined by Federal Law (i.e., need for contextual appraisals to derive intervention information); (b) response classes for education often cut across diagnostic categories (i.e., problem behavior can be functionally related to instructional variables, for example, 'going to school behaviors' are functionally related to behavioral and academic goals); (c) measurement designs can be similar for academic and social responses; (d) intervention designs also may have many similarities; and (e) problem solving unifies the approaches "organized around what a child needs to learn" (Deno & Gross, 1973, p. 104).

A three level integrated model is described by O'Shaughnessy, Lane, Gresham, and Beebe-Frankenberger (2003). The first level, described as *Benchmark intervention*, is preventative, for all children, and directed at learning and behavior. Examples of intervention include early literacy instruction, social skills, and conflict resolution. A second level, referred to

as *strategic intervention*, targets children described as "at-risk for learning and/or behavior problems" (p. 31). The third level is called *intensive intervention* and may include small group or individualized instruction, special education, and community involvement.

Response to Intervention with Young Children: *Key Concepts*

Current models have evolved for school-aged children, but a considerable body of research supports key components of RTI with young children with targets of intervention that include social behaviors and a wide range of developmental skills (Barnett, Bell, & Carey, 1999; Webster-Stratton, 1999). The major components are the same as those presented earlier: (a) universal screening and assessment on significant developmental indicators and environmental/instructional adequacy (McConnell, 2000); (b) relatively brief, well planned, discrete intervention (or instructional trials) that are hierarchically (i.e., Brown, Odom, & Conroy, 2001) and/or sequentially organized by levels of intensity; and (c) continuous progress monitoring, with outcomes evaluated by single subject design facets. Intervention choice is guided by empirical support and judgments of 'positiveness' and 'naturalness' and change agents frequently include parents, teachers, as well as children's peers (Dunst, Bruder, Trivette, Hamby, Raab, & McLean, 2001; Kohler, Anthony, Steighner, & Hoyson, 2001).

What remains to make the model viable is organizing RTI research and concepts into a service delivery system, field-testing the model and associated decision rules, and evaluation. This would include an agenda of staff development, a strong set of procedures including criteria for technical adequacy, and exemplars for assessment and intervention. In the next section, a tiered model of services and permanent products for procedures and evaluation are reviewed.

Rationale for a Permanent Products Approach

Drawing on other models and state-funded research, significant characteristics of a preschool RTI model include *permanent products* for documentation. The first step in developing a permanent products approach is establishing written procedures. Related to basic concepts in intervention research (e.g., Billingsley, White, & Munson, 1980; Gresham, 1989), *procedural checklists* have been used more broadly for describing and documentation key events in service delivery. To build a procedural checklist, critical features of a model are described as a series of professional practice steps defining the model (Thomas, Bastien, Stuebe, Bronson, & Yaffe, 1987). The steps for the integrated preschool RTI model were derived from the intervention 'practices' literature and empirically refined throughout research and training (Barnett, Daly, et al., 1999; Barnett, Pepiton, et al., 1999; Lentz, Allen, & Ehrhardt, 1996).

A procedural checklist serves three significant functions. First, it clarifies, targets, and facilitates monitoring of intervention development, technical and legal 'checks,' and accountability procedures by guiding system-level implementation and the delineation of the specific building-level, team-level, and child-level responsibilities and outcomes. Second, it adds legal and ethical support by providing an 'informed consent' communication to parents, schools, agencies, and teachers about professional actions. Third, it results in key 'permanent products' that may be used for team progress monitoring for the process of problem-solving as well as eligibility determination, as needed, and may also serve as the replicated variable in research for model validation.

As examples, permanent products would include evidence of hypothesis formation, progress monitoring in the form of graphed data, intervention plans, and decision rule use.

Research supports the use of specific intervention plans in the form of simple scripts supported with training, feedback, and strong agency sanctions (Ehrhardt, Barnett, Lentz, Stollar, & Reifin, 1996; Martens, Hiralall, & Bradley,1997; Noell, Witt, LaFleur, Mortenson, Ranier & LeVelle, 2000; Sterling-Turner, Watson, & Moore, 2002). The most serious questions, that of eligibility for specialized service decisions, are based on documented progress (permanent products) of a child's cumulative responses to intervention through 3 tiers of progress monitoring, not intra-individual ability discrepancies derived from the administration of one-point-in-time tests.

A Response to Intervention Model for Preschool Services

For the most part, response to intervention has evolved to address specific learning disabilities due to high rates of unproductive classification as well as separate models for highly challenging behaviors (c.f. O'Shaughnessy, et al., 2003). We offer an integrated model that would accommodate agency agendas that include the adequacy of instructional environments and an extended range of disability services.

Common Elements Through 3 Tiers

Adequacy of instructional environments as a foundation. A major concept found in *IDEA* and *No Child Left Behind* is that decisions are based on child characteristics contextualized by the capacity of a teacher and classroom to meet the needs of children. This capacity entails both instructional practices to effectively attain learning outcomes for all students as well as a positive, preventive system of management and support for social and behavioral competence. At the building level, effective schools may be characterized by validated instructional practices, extended school day/year policies, child advocacy programs, and proactive school-wide plans for

teaching and encouraging appropriate social behaviors to support child outcomes.

Increasing and decreasing intensity intervention decisions. Increasing intensity designs are based on discrete, sequential intervention trials ordered on a continuum that *builds* in intensity to refine intervention programs to optimal levels. *Decreasing intensity* designs start with more comprehensive or multi-component interventions in response to judgments of high risk. Intervention facets are systematically withdrawn to allow interventions to become more natural, self-regulated, and ecologically sustainable. The ongoing objective for both in RTI is applying the least amount of intervention to accomplish objectives within typical educational environments (Barnett, et al., in press).

Problem solving, evidenced-based intervention, teacher and parent supports. Common elements across the tiers include: (a) measuring student outcomes, (b) setting goals and developing step-by-step intervention plans, and (c) using graphic analysis of intervention results and ongoing comparisons of the student's performance with baseline data and goals. Systematic, problem-solving approaches are characterized by evidence-based interventions and by data-based decision making. System level supports for problem-solving and team consultation also are programmed (Crone & Horner, 1999-2000; Ikeda, Tilly, Stumme, Volmer, & Allison, 1996). Evidence for these supports include policy statements for granting time off for teachers to participate in problem-solving activities, parent outreach such as sponsored babysitting services to facilitate parental involvement, allocation of resources (e.g., substitute teachers, funds, internet access) to support team activities, and public recognition of successful service plans displayed in school message boards, newsletters, and student handbooks.

Overview of technical adequacy. While linked to traditional reliability and validity

concepts, the basics of RTI technical adequacy require different practices: (a) demonstrating reasonableness of measurement selection leading to the accurate and socially valid description of a functional discrepancy between children's current performance and meaningful goals, (b) showing defensible intervention selection (i.e., empirically defensible interventions in a sound sequence), (c) documenting intervention adherence, and (d) measuring intervention outcome.

By adding RTI, technical adequacy would include clarification of decision rules for (e) interpreting levels of intensity and (f) changing intervention plans (within and between tiers). Decision rule use is a topic of considerable complexity but at the least, data would include goal setting, repeated measurement and graphed lines of progress towards goals, evaluating level and trend of a discrete instructional (or behavioral) 'trial,' and applying 'rules' to modify instruction (or intervention) or change goals (i.e., Fuchs & Fuchs, 1986; Haring, Liberty, & White, 1980; Howell & Nolet, 2000; Wolery, Bailey, & Sugai, 1988).

Defining 'Tiers' of Preschool Services

Tier 1: class-wide. At the first tier, class-wide interventions address curriculum, activities, and routines as well as teacher instructional or managerial practices that positively influence rate of learning and child engagement for all children, and reduce overall rates of disruptive behaviors (e.g., Brown, et al., 2001; Hemmeter, 2000; Joseph & Strain, 2003; VanDerHeyden, Witt, & Gatti, 2001). This Tier includes teachers' efforts to supplement core programs as needed (including interventions, modifications, adaptations) and manage positive behavior support programs for classes. Class-wide agency supports would include helping a teacher with regard to successfully implementing a scientifically-based curriculum for *all*

children in the classroom and positive behavior supports, and managing related classroom routines.

Tier 1 may include class-wide interventions to increase engagement or changes, modifications, adaptations, and interventions for individual students if necessary. The defining characteristic for Tier 1 is that resources necessary for these changes are managed by the classroom teacher. For an individual child, the key objectives would be to determine the effects of a class-wide intervention. The teacher can engage in problem solving alone or in collaboration with consulting professionals and/or parents. These consultations would follow the problem-solving model, be linked to research-based interventions, and would include progress monitoring for the individual child as needed.

Tier 1 parent support would include comprehensive programming and orientations for parents (McWilliams, Tocci, & Harbin, 1998), as well as group programs for parents on educationally related topics such as promoting early literacy and language as well as positive social behaviors (e.g., Dunst, et al, 2001; McNeill & Fowler, 1999; Webster-Stratton, 1999). Programs usually would be implemented collaboratively by teachers and agency personnel.

Tier 2: Embedded or small group. This tier includes *agency and classroom-based efforts* to improve outcomes for children not responding satisfactorily to Tier 1 efforts. It includes interventions that can be integrated into classroom routines with supportive assistance for teacher management from professionals. Embedded or small group interventions have both been found to be effective when targeting a diverse array of learning needs (Wolery, 2000; Wolery, Anthony, Caldwell, Snyder, & Morgante, 2002). Embedded interventions provide increased practice opportunities for specific skills, such as making functional requests (Tirapelle & Cipani,

1992) within a class routine (i.e., art, tooth brushing).

Tier 2 also may involve a range of specialized curricular programs, with assistance for teacher responsibility or outside of teacher management, for groups of children. Major examples include social skills programs, supplemental language interventions and cross-age or volunteer tutors for children. These group programs traditionally are referred to as ‘remedial.’ Specialized professionals may help teachers plan and monitor embedded interventions or may help implement group interventions. Multiple professionals (including teachers) may engage in the decisions to begin a supplemental program, monitor progress while in the program, modify program procedures or attempt additional interventions, and judge outcomes/next steps. Embedded or supplemental programs are scientifically-supported as effective for the class of concerning performances/problems and are linked to Tier 1 core curricula and positive behavior support. Parents are informed about Tier 2 interventions and have input into goal setting and other decisions. Decisions for any Tier 2 program are based on more frequent (than Tier 1) data from school- or class-wide progress monitoring systems. Agencies develop objective ‘rules’ for making decisions about: returning to Tier 1, making changes in Tier 2 methods, or moving to Tier 3.

Tier 2 parent support is centered on giving parents additional choices for involvement building on Tier 1 programs. Tier 2 parent roles may include more frequent progress reports and additional supportive activities for Tier 2 interventions (e.g., Hancock, Kaiser, & Delaney, 2002).

Tier 3: Intensive and individualized. Tier 3 is characterized by frequent progress

monitoring with at least school-wide data, structured problem solving and more idiosyncratic empirical interventions, and data-based decision making for children not responding to interventions in Tier 2. Tier 3 interventions are still based at least initially in a child's classroom. An expanded team of parent(s) and professionals develop more intensified intervention for children not responding satisfactorily to Tier 2 efforts, and child and teacher supports are added (Wolery, Brashers, & Neitzel, 2002). An example of a Tier 3 intervention linked to Tiers 1 and 2 is an individualized (or activity) schedule to provide additional structure for the child and teachers/support persons, facilitate transitions, and embed instructional objectives and reward systems (Meisbov, Browder, & Kirkland, 2002).

Tier 3 parental roles include decision making and further choices about level of involvement, which may be considerable and may build on Tier 1 and 2 interventions (i.e., Dunst et a., 2001; Hancock et al., 2002). One role for participation is functional assessment to further refine intervention design for Tier 3 (O'Neill, et al., 1997). As another example, individualized (activity) schedules may be used across home and school settings to target practice and generalization for more serious language and behavior challenges (Mesibov, et al., 2002).

Tier 3: Eligibility. In spite of the complexities, response to intervention methods meet the objectives of legislation intended to improve academic responding and disability services. Rather than creating a 4th tier, we integrate eligibility into Tier 3 to emphasize the importance of timing and flexibility of specialized services linked to typical environments and not a 'place' for challenging children. The foundations of eligibility are a technically adequate data set through 3 tiers of intervention.

A Closer Look at Technical Adequacy Across 3 Tiers

Technical adequacy is documented by the permanent products of problem solving. Table 1 describes specific technical checks for each tier of services. The table is daunting because it addresses many classes of reform, and many forms of technical adequacy. Alternatively, it represents a different lexicon for technical adequacy, and judgments of relative complexity would take into account comparisons with the complexities of traditional technical adequacy (e.g., Wainer & Braun, 1988). Also, much has been published recently about what is meant by scientific evidence on a continuum of confidence (c.f., Kratochwill & Stoiber, 2000). We have tried to stay true to these ideals but with two differences. Practice in a 3 tier model also includes traditional criteria for intervention confidence, emphasizing: (a) internal validity or the degree of evidence that may exist that the intervention was responsible for change, even on a relatively small scale (published single subject designs with within and across series replications); or (b) an intervention developed in consultation with a teacher that is based on a functional assessment and replicated *principles* of learning (opportunities to practice, positive attention). The key idea is communicating accurate *confidence* in the intervention to consumers of services.

All tiers include the technical adequacy of problem solving (Macmann et al., 1996). Tier 1 includes the 'validity' or scientific evidence that may be marshaled for curriculum (e.g., Engelmann & Osborn, 1997; Webster-Stratton, Reid & Hammond, 2001), large scale assessments and progress monitoring (Fuchs & Fuchs, 1986; McConnell, 2000; VanDerHeyden, et al., 2001), and environmental measures (Greenwood, Carta, & Dawson, 2000; Harms, Clifford, & Cryer, 1998; Nichols & Barnett, 2004; Smith & Dickinson, 2002). Various environmental measures of classrooms have different purposes and methods. For example, the

observation system by Nichols and Barnett (2004) uses interval recording of instructional and managerial contacts, interactive reading, conversations, incidental teaching, positive attention, and momentary time scan sampling of child engagement as a basis for feedback and support for teachers.

Tiers 2 and 3 emphasize specific intervention outcome studies with the difference being that they are hierarchically arranged in intensity. In Tier 3, when addressing eligibility, technical adequacy is based on (a) a task analysis of idiosyncratic interventions (LeLaurin & Wolery, 1992); (b) the reliability and validity of goal, trend, and slope estimates for evaluating the discrepancy between educational performance of the referred child in comparison to peer norms (continuing from Tiers 1 and 2), *and* (c) comparisons of intensity to peer norms for a desired change in performance that is not responsive to iterations of problem-solving. Judgments are made by considering a data set across 3 tiers.

Putting the components together. A 3-tier sequence provides a cumulative intervention history for a child ordered by intervention intensity. Figure 1 shows an abbreviated version of the permanent products checklist to address procedural reliability. Key events that are used to track a child's progression through 3 tiers of services for eligibility decision-making are included. Figure 2 shows a hypothetical data set for a preschool class and progression for a Tier 1 intervention, as well as a group intervention (Tier 2) and individualized intervention (Tier 3). Actual figures for classrooms and individual children would vary by targeted measures, and classrooms and children may have multiple figures. Data in Tier 2 could also be group data such as median plots by response class of children receiving the same intervention (social skills curriculum). A similar data set could illustrate any measurement of a 'keystone' outcome such as

early literacy or social skill. Table 2 includes evidence-based interventions that have been used at Head Starts. The cumulative intervention results for a child may be used to help with judgments concerning special education eligibility.

Case Contributions to Team Fluency and Capacity Building

An objective of applied behavior analysis and a full model of intervention-based services through RTI is that outcomes are used to further the capacity of agencies to meet future challenges (Baer, Wolf, & Risley, 1987). Capacity building is evidenced by school improvement plans that include the systematic use of data based on actual referrals to improve educational supports for all students (e.g., Anderson, Russo, Dunlap, & Albin, 1996; O'Shaughnessy et al., 2003). In other words, capacity building includes action plans that are linked to the base rates of agency case characteristics. An agenda for building team fluency (Binder, 1996) and capacity to meet challenges can be created by considering the children served, the referrals for more intensive intervention services, the nature of needed supports, the outcomes of interventions, and the qualities of effective interventions, in order to look systematically at accomplishments and challenges to service delivery programs. As an example, targeted variables in the data base for Ohio's Early Childhood Education (ECE) Intervention project (Barnett et al., 1997; Barnett, Pepiton et al., 1999) included referrals for supporting diverse learners in general classrooms (e.g., success with typical routines), improving child competence in social play, increasing engagement in learning activities, and improving communication, literacy, and other pre-academic skills. These referrals and targeted variables were used as ongoing checks for building the capacity of the ECE Intervention project. Capacity building included making certain that (a)

consultants were adequately prepared to help guide problem-solving; and (b) parents, teachers, and teams were well supported in implementing interventions for individual target students as well as for classes or groups of children. Key checks were used to ensure that measurement designs were appropriate to the referrals and targeted variables and that teams considered the strongest possible range of positive empirical interventions. Measures included ways to accurately, efficiently, and sensitively monitor performance and behavior (e.g., curriculum-based measures; time sampling for engagement or disruptive behaviors) and intervention use. Evaluation of system-wide effects can be based on evidence of use of effective classroom management strategies and instructional practices derived from outcome data from targeted interventions across students served.

Discussion and Conclusions

Challenges, Criticism, and Limitations

Highly motivated, well trained professionals and a dedicated whole-school context are two of the components needed for an RTI model to work effectively. A RTI model needs to be well defined, operationalized, and carried out with procedural fidelity. Research demonstrates the need for support for teachers to implement intervention practices (Jones, Wickstrom, & Friman, 1997; Kohler, et al., 2001; McBride & Schwartz, 2003).

Criticism can be made about the complexity of technical adequacy of methods underlying RTI. There are many points of analysis, but central are the adequacy of instructional environments, measurement of key variables at 3 tiers, qualities of decision rules that are applied to data, accuracy of interpretations, validity of intervention design and sequence, and judgments that must occur about intensity linked to eligibility. These concerns have some type of

counterpoint in traditional diagnosis and classification with two very substantial differences: (a) quantified contextual patterns and (b) an iterative process of brief, sequenced, discrete interventions. These two elements in combination with the supports and adjustments to intervention plans gives RTI its potential strength over one-point-in-time approaches.

A key challenge to implementation is identifying adequate dynamic indicators of progress that are linked to important functional outcomes and identifying which instructional strategies are most powerful at effecting desired change (Rule, Losardo, Dinnebeil, Kaiser, & Rowland, 1998). Whereas goals have been operationalized in an idiographic fashion in multiple single-subject studies, fewer investigations have sought to create and study measures that might be used to track the general outcomes resulting from intervention. Nevertheless, several themes have emerged that warrant consideration as meaningful dynamic indicators for use with young children.

For example, studies with elementary-aged children have found engagement to be an important indicator of academic success (Greenwood,1991) that can be improved through intervention such as the provision of peer tutoring (DuPaul , Ervin, Hook, & McGoey, 1998) or peer support (Cushing & Kennedy, 1997); that is controlled by the reinforcement schedules in effect (Martens, Lochner, & Kelly, 1992); and is sensitive to environmental arrangements (Logan, Bakeman, & Keefe, 1997). Many early intervention researchers have suggested that engagement is similarly an important indicator for young children and have demonstrated its reliable measurement and sensitivity to certain environmental conditions (Connell, Carta, & Baer, 1993; McWilliam, Trivette, & Dunst, 1985; Dunlap, et al., 1994; Reinhartsen, Garfinkle, &

Wolery, 2002; Risley & Cataldo, 1974).

Some have proposed that engagement is a necessary foundational skill upon which other critical skills are built (Reinhartsen, et al., 2002). Critically, children who are not engaged with the materials and people in their learning environments are not likely to derive maximal benefit from that setting because they miss important learning and practice opportunities. Beginning with the earliest interactions, it seems that skills beget additional skills and in a faster and cumulative fashion (Brophy & Good, 1970; "Mathew Effect"; Stanovich, 1986). For example, children who have fluent social interaction skills are likely to encounter (as a product of their behavior) additional opportunities to practice social interaction skills and receive reinforcement at higher rates for demonstrating those skills relative to children with weak social interaction skills. Language skills such as vocalizations and verbalizations per unit of time are potentially promising dynamic indicators (Hart & Risley, 1995, 1999). Hart and Risley (1995) found that children who had more experience with language (were spoken to more frequently), received interaction with their parents that was characterized as responsive (i.e., following the child's attentional lead), and received a greater number of encouraging versus discouraging statements had a more rapidly growing vocabulary relative to children with less language experience. An analogous pattern is apparent with what might be described as "ready to learn" behaviors. Patterson and colleagues found in a series of studies that a particular type of parent-child interaction termed "coercive interaction" established aggressive and noncompliant behaviors at a young age that persisted over time and were linked to multiple poor outcomes (e.g., school failure) without intervention (Patterson, 1986; Kazdin, 1987). Early literacy skills follow the same pattern. Children with weak prerequisite skills (often related to language experience/skill;

Torgeson et al., 2001) have weaker reading skills relative to their peers and grow at a slower pace without intervention. Given early, effective intervention, deficits can be repaired and future deficits can be prevented (Lennon & Slesinski, 1999; Torgeson, 2002).

With older children, curriculum-based measurement is documented as a tool that can be used to obtain reliable indicators of child performance, that can be used formatively to enhance instruction, and that reflects functional and meaningful outcomes. A group of researchers has been working to build measures that might be similarly used with young children. Good and Kaminski (1996) and Kaminski and Good (1996) developed a set of measures to assess phonemic awareness in young children (Dynamic Indicators of Early Literacy Skills) and established benchmarks for identifying children at risk for reading failure (Good, Simmons, & Kame'enui, 2001). These measures can be used for universal screening to determine which students need assistance, to track the effects of the assistance provided, and to evaluate the outcome of the intervention effort for children in kindergarten and potentially in four-year old preschool classrooms. With even younger children, promising progress has been made toward identifying indicators of language development (Hart & Risley, 1995, 1999; Luze, et al., 2001), social interaction (Carta, Greenwood, Luze, Cline, & Kuntz, 2004), and motor development (Greenwood, Luze, Cline, Kuntz, & Leitschuh, 2002). These measures and others like them merit further scrutiny to determine their utility for tracking and guiding intervention efforts and judging outcomes in early intervention.

Perhaps because adequate dependent measures have not been universally accepted and consistently defined, the independent variable (i.e., intervention) has been difficult to quantify.

Studies have found that teachers frequently use strategies to engage children in interaction in the classroom, but rarely systematically apply these instructional strategies in order of increasing intensity to prompt or promote the occurrence of higher-level targeted skills (Roberts, Bailey, & Nychka, 1991; Schwartz, Carta, & Grant, 1996). The use of these systematic instructional strategies has been linked to better outcomes for participants (Chiara, Schuster, Bell, & Wolery, 1995; Losardo & Bricker, 1994; McBride & Schwartz, 2003; McGee, Krantz, & McClannahan, 1985; Schwartz, Carta, & Grant, 1996). Practitioners using an RTI model in early intervention and preschool settings will face the dual challenge of measuring the degree to which relevant instruction effectively occurs (Rule, et al., 1998), and training teachers to ensure adequate integrity with change efforts to ensure improved outcomes for children (Holcombe, Wolery, & Snyder, 1994).

Another criticism is that of potential cost. However, studies show that behavioral problem solving may take about the same amount of time as so-called thorough testing (Barnett, Pepiton, et al., 1999) but problem solving is highly dependent on team fluency and less dependent on child characteristics. A key ingredient in successful reform is producing effective outcomes in a cost-reasonable manner. Costs are minimized by (a) improving instructional ecologies to reduce the number of referrals and negative outcomes for students (learning and behavioral), (b) using universal screening and information already available from parents and teachers, (c) adding measures with high potential yield for usefulness *only as needed*, and (d) giving up unhelpful strategies. If the critics are right regarding traditional child measurement for disability services, the savings may be considerable.

Perhaps the greatest challenge is that a full 3-tier model requires a school culture

committed to scientific advances in education and services. As reviewed, current models and outcome data exist for accountability, decision making for intervention intensity and specialized services, and technical adequacy.

Contributions

Educational reform has proven to be vulnerable as a construct and rallying cry. The cumulative and systematic growth of educational innovation based on outcomes has not taken place as expected (Fuchs, Fuchs, Harris, & Roberts 1996). There are many points of analysis when considering educational reform. We emphasized change that is needed in what is measured, what is documented about the change process, and how this information is used. RTI is a logical and empirical way to organize services and create a data set for potentially sound decisions about special services. We have focused on permanent products that communicate for each case the quality of intervention-based services in place when milestone educational decisions are made.

Assessments of child performance leading to a functional discrepancy analysis and showing needed interventions are currently achievable. Decisions about special services can be derived from data-based descriptions of interventions and outcomes demonstrating needed support for children and teachers.

References

AERA, APA, & NCME (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education). (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Alfonso, V. C., Oakland, T. D., LaRocca, R., & Spanakos, A. (2000). The course on individual cognitive assessment. *School Psychology Review, 29*, 52-64.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.) Upper Saddle River, NJ: Prentice Hall.

Anderson, J. L., Russo, A., Dunlap, G., & Albin, R. W. (1996). A team training model for building the capacity to provide positive behavioral supports in inclusive settings. In L. K. Koegel, R. L. Koegel, & G. Dunlap, (Eds.), *Positive behavioral support: Including people with difficult behavior in the community* (pp. 467-490). Baltimore, MD: Brookes

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20,* 313-327.

Barnett, D. W. (1988). Professional judgment: A critical appraisal. *School Psychology Review, 17,* 656-670.

Barnett, D. W., Bell, S. H., Bauer, A., Lentz, F. E. Jr., Petrelli, S., Air, A., & Hannum, L. (1997). The early childhood intervention project: Building capacity for service delivery. *School Psychology Quarterly, 12,* 293-315.

Barnett, D. W., Bell, S. H., Gilkey, C. M., Lentz, F. E., Jr., Graden, J. L., Stone, C. M., & Smith, J. J. (1999). The promise of meaningful eligibility determination: Functional

intervention-based multifactored preschool evaluation. *Journal of Special Education, 33,* 112-124.

Barnett, D. W., Carey, K. T., & Bell, S. (1999). *Designing preschool interventions.* New York: Guilford.

Barnett, D. W., Daly, E. J. III, Hampshire, E. M., Hines, N. R., Maples, K. A., Ostrom, J. K., & Van Buren, A. E. (1999). Meeting performance-based training demands: Accountability in an intervention-based practicum. *School Psychology Quarterly.*

Barnett, D. W., Daly, E. J. III, Jones, L. M., & Lentz, F. E. Jr. (in press). Response to intervention: Decisions from increasing and decreasing intensity single case designs. *The Journal of Special Education*.

Barnett, D. W., Lentz, F. E., Jr., & Macmann, G. M. (2000). Psychometric qualities of professional practice. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral Assessment* (2nd ed., pp. 355-386). NY: Guilford.

Barnett, D. W., Macmann, G. M., & Carey, K. T. (1992). Early intervention and the assessment of developmental skills: Challenges and directions. *Topics in Early Childhood Special Education, 12,* 21-43.

Barnett, D. W., Pepiton, A. E., Bell, S. H., Gilkey, C. M., Smith, J. J., Stone, C. M., Nelson, K. I., Maples, K. A., Helenbrook, K., & Hannum, L. E. (1999). Evaluating early intervention: Accountability methods for service delivery innovations. *Journal of Special Education, 33,* 177-188.

Bell, S. H., & Barnett, D. W. (1999). Peer micronorms in the assessment of young children:

Methodological review and examples. *Topics in Early Childhood Special Education, 19,* 112-122.

Billingsley, F., White, D. R., & Munson, R. (1980). Procedural reliability: A rationale and an example. *Behavioral Assessment, 2,* 247-256.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19,* 163-197.

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61,* 365-374.

Brown, W. H., Odom, S. L., & Conroy, M. A. (2001). An intervention hierarchy for promoting young children's peer interactions in natural environments. *Topics in Early Childhood Special Education, 21,* 162-175.

Carr, E. G., Horner, R. H., & Turnbull, A. P. (1999). *Positive behavior support for people with developmental disabilities: A research synthesis.* Washington, DC: American Association on Mental Retardation.

Carta, J.J., Greenwood, C.R., Luze, G.J., Cline, G., & Kuntz, S. (2004). Developing a general outcome measure of growth in social skills for infants and toddlers. *Journal of Early Intervention, 26,* 91-114.

Carta, J. J., Schwartz, I. S., Atwater, J. B., & McConnell, S. R. (1991). Developmentally appropriate practice: Appraising its usefulness for young children with disabilities. *Topics in Early Childhood Special Education, 11,* 1-20.

Chiara, L., Schuster, J. W., Bell, J. K., & Wolery, M. (1995). Small-group massed-trial and

individually-distributed-trial instruction with preschoolers. *Journal of Early Intervention, 19*, 203-217.

Colvin, G., Kameenui, E. J., & Sugai, G. (1993). Reconceptualizing behavior management and school-wide discipline in general education. *Education and Treatment of Children, 16,* 361-381.

Connell, M. C., Carta, J. J., & Baer, D. M. (1993). Programming generalization of in-class transition skills: Teaching preschoolers with developmental delays to self-assess and recruit contingent teacher praise. *Journal of Applied Behavior Analysis,* 26, 345-352.

Crone, D. A., & Horner, R. H. (1999-2000). Contextual, conceptual, and empirical foundations of functional behavioral assessment in schools. *Exceptionality, 8,* 161-172.

Cushing, L. S., & Kennedy, C. H. (1997). Academic effects of providing peer support in general education classrooms on students without disabilities. *Journal of Applied Behavior Analysis,* 30, 139-150.

Deno, S. & Gross, J. (1973). The Seward-University Project: A cooperative effort to improve school services and university training. In E. N. Deno (Ed.), *Instructional alternatives for exceptional children* (pp. 104-121). Arlington, VA: CEC.

Dunlap, G., dePerczel, M., Clarke, S., Wilson, D., Wright, S., White, R., & Gomez, A. (1994). Choice making to promote adaptive behavior for students with emotional and behavioral challenges. *Journal of Applied Behavior Analysis, 27*, 505-518.

Dunst, C. J., Bruder, M. B., Trivette, C. M., Hamby, D., Raab, M., & McLean, M. (2001). Characteristics and consequences of everyday natural learning opportunities. *Topics in*

Early Childhood Special Education, 21, 68-92.

DuPaul, G. J., Ervin, R. A., Hook, C. L., & McGoey, K. E. (1998). Peer tutoring for children with attention deficit hyperactivity disorder: Effects on classroom behavior and academic performance. *Journal of Applied Behavior Analysis, 31,* 579-592.

Ehrhardt, K. E., Barnett, D. W., Lentz, F. E., Jr., Stollar, S. A., & Reifin, L. H. (1996). Innovative methodology in ecological consultation: Use of scripts to promote treatment acceptability and integrity. *School Psychology Quarterly, 11,* 149-168.

Engelmann, S. & Osborne, J. (1997). *Language for learning.* SRA.

Fiske, D. W. (1982). Convergent-discriminant validation in measurements and research strategies. *New directions for methodology of social and behavioral science, 12,* 77-92.

Fletcher, J. M., Lyon, R. G., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., Shaywitz, S. F., & Shaywitz, B. A. (2001). *Classification of learning disabilities: An evidence-based evaluation.* (U.S. DOE White Paper).

Flugum, K. R., & Reschly, D. J. (1994). Prereferral interventions: Quality indices and outcomes. *Journal of School Psychology, 32,* 1-14.

Fuchs, D., Fuchs, L. S., Harris, A. H., & Roberts, P. H. (1996). Bridging the research-to-practice gap with mainstream assistance teams: A cautionary tale. *School Psychology Quarterly, 11,* 244-266.

Fuchs. L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199-208.

Fuchs. L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13,*

204-219.

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disabilities Quarterly, 25*, 33-44.

Galagan, J. E. (1985). Psychoeducational testing: Turn out the lights, the party's over. *Exceptional Children, 52,* 288-299.

Gersten, R. M., Becker, W. C., Heiry, T. J., & White, W. T. (1984). Entry IQ and yearly academic growth of children in direct instruction programs: A longitudinal study of low SES children. Educational Evaluation and Policy Analysis, 6, 109-121.

Gettinger, M., & Stoiber, K. C. (1999). Excellence in teaching: Review of instructional and environmental variables. In C. R. Reynolds and T.B. Gutkin (Eds.), *Handbook of School Psychology* (3rd ed., pp. 933-958). New York: Wiley.

Good, R. H., Kame'enui, E., Simmons, D., & Chard, D. (n.d.). *Using Dynamic Indicators of Basic Early Literacy Skills in a Schoolwide model for primary, secondary, and tertiary prevention. University of Oregon*. Unpublished manuscript.

Good, R. H., Simmons, D., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

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., III, Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic

Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.) *Best practices in school psychology* (4th ed.; pp. 679-700). Bethesda, MD: National Association of School Psychologists.

Greenwood, C. R. (1991). Longitudinal analysis of time, engagement, and achievement in at-risk versus non-risk students. *Exceptional Children*, 57, 521-535.

Greenwood, C. R., Carta, J. J., & Dawson, H. (2000). Ecobehavioral assessment systems software (EBASS): A system for observing in education settings. In T. Thompson, D. Felce, & F. J. Symons (Eds.), *Behavioral observation: Technology and applications in developmental disabilities* (pp. 229-251). Baltimore, MD: Brookes.

Greenwood, C.R., Luze, G.J., Cline, G., Kuntz, S., & Leitschuh, C. (2002). Developing a general outcome measure of growth in movement for infants and toddlers. *Topics in Early Childhood Special Education, 22,* 143-157.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation/prefererral intervention. *School Psychology Review, 18,* 37-50

. Gresham, F. M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. *School Psychology Review, 20,* 20-36.

Gresham, F. M. (2001). *Responsiveness to intervention: An alternative approach to the identification of learning disabilities (2nd draft).* Paper presented for the OSEP Learning Disabilities Initiative, Office of Special Education Programs, U. S. Department of Education, Washington, DC.

Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1998). Agreement between school study team decisions and authoritative definitions in classification of students at-risk for mild

disabilities. *School Psychology Quarterly, 13,* 181-191.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly, 16,* 31-55.

Hancock, T. B., Kaiser, A. P., & Delaney, E. M. (2002). Teaching parents of preschoolers at high risk: Strategies to support language and positive behavior. *Topics in Early Childhood Special Education, 22,* 191-212.

Haring, N. G., Liberty, K. A., & White, O. R., (1980). Rules for data-based decision strategy decisions in instructional programs. In W. Sailor, B. Wilcox, & L. Brown (Eds.), *Methods of instruction for severely handicapped students* (pp. 159-192). Baltimore: Paul Brookes.

Harms, T. & Clifford, R. M. & Cryer, D. (1998). *Early childhood environment rating scale (rev. ed).* New York: Teachers College Press.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes Publishing.

Hart, B., & Risley, T. R. (1999). *The social world of children learning to talk*. Baltimore, MD: Paul H. Brookes Publishing.

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.

Hemmeter, M. L. (2000). Classroom-based interventions: Evaluating the past and looking toward

the future. *Topics in Early Childhood Special Education, 20,* 56-61.

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4,* 49-73.

Horner, R. H., & Sugai, G. (2000). School-wide positive behavioral support: An emerging initiative. *Journal of Positive behavioral Interventions, 2,* 231-232.

Howell, K. W., & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making.* Belmont, CA: Wadsworth.

IDEA. *Individuals with Disabilities education Act Amendments of 1997* (PL 105-17). 20 U.S.C. Chapter 33, Sections 1400-1485 (Statute).

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.

Jones, K. M., Wickstrom, K. F., & Friman, P. C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12,* 316-326.

Joseph, G. E. & Strain, P. S. (2003). Comprehensive evidence-based social-emotional curricula for young children: An analysis. *Topics in Early Childhood Special Education, 23,* 65-76.

Kaminski, R. A., & Good III, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215-227.

Kazdin, A. E. (1987). Treatment of antisocial behavior in children: Current status and future directions. *Psychological Bulletin, 102,* 187-203.

Keith, T. Z., Krantzler, J. H., & Flanagan, D. P. (2001). What does the Cognitive Assessment

System measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson Tests of Cognitive Ability (3rd ed.). *School Psychology Review, 30,* 89-119.

Kohler, F. W., Anthony, L. J., Steighner, S. A., & Hoyson, M. (2001). Teaching social interaction skills in the integrated preschool: An examination of naturalistic techniques. *Topics in Early Childhood Special Education, 21, 93-103.*

Kohn, A. (2001). Five reasons to stop saying, "good job!" *Young Children, 40 (3),* 3-9.

Kratochwill, T. R., & Stoiber, K. C. (2000). Empirically supported interventions and school psychology: Rationale and methodological issues–Part II. *School Psychology Quarterly, 15*, 233-253.

Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for white and black students from the WISC-III standardization sample and for black students referred for psychological evaluation. *School Psychology Review, 30,* 70-88.

LeLaurin K., & Wolery, M. (1992). Research standards in early intervention: Defining, describing, and measuring the independent variable. *Journal of Early Intervention, 16,* 275-287.

Lennon, J.E. & Slesinski**,** C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review, 28,* 353-364.

Lentz, F. E., Jr., Allen, S. J., & Ehrhardt, K. E. (1996). The conceptual elements of strong interventions in school settings. *School Psychology Quarterly, 11,* 118-136.

Linehan, S. A., Brady, M. P., Hwang, C. (1991). Ecological versus developmental assessment:

Influences on instructional expectations. *Journal for the Association of Persons with Severe Handicaps, 16,* 146-153.

Logan, K. R., Bakeman, R., & Keefe, E. B. (1997). Effects of instructional variables on engaged behavior of students with disabilities general education classrooms. *Exceptional Children, 63,* 481-497.

Losardo, A., & Bricker, D. (1994). Activity-based intervention and direct instruction: A comparison study. *American Journal on Mental Retardation, 98,* 744-765.

Lorimer, P. A., Simpson, R. L., Myles, B. S., & Ganz, J. B. (2002). The use of social stories as a preventative behavioral intervention in a home setting with a child with autism. *Journal of Positive Behavioral Interventions, 4,* 53-60.

Luze, G. J., Linebarger, D. L., Greenwood, C. R., Carta, J. J., Walker, D., Leitschuh, C., & Atwater, J. B. (2001). Developing a general outcome measure of growth in expressive communication of infants and toddlers. *School Psychology Review, 30,* 383-406.

Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the Verbal-Performance dichotomy of the Weschler Scales. *School Psychology Quarterly, 9,* 161-197.

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "Intelligent Testing" approach to the WISC-III. *School Psychology Quarterly, 12,* 197-234.

Macmann, G., & Barnett, D. (1999). Diagnostic decision making in school psychology: Understanding and coping with uncertainty. In C. R. Reynolds and T.B. Gutkin (Eds.), *Handbook of School Psychology* (3rd ed., pp. 519-548). New York: Wiley.

Macmann, G. M., Barnett, D. W., Allen, S. J., Bramlett, R. K., Hall, J. D., & Ehrhardt, K. E. (1996). Problem solving and intervention design: Guidelines for the evaluation of technical adequacy. *School Psychology Quarterly, 11,* 137-148.

Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of classification agreement. *Journal of Special Education, 23,* 127-149.

Martens, B. K., Bradley, T. A., & Eckert, T. L. (1997). Effects of reinforcement history and instructions on the persistence of student engagement. *Journal of Applied Behavior Analysis, 30,* 569-572.

Martens, B. K., Eckert, T. L., Bradley, T. A., & Ardoin, S. P. (1999). Identifying effective treatments from a brief experimental analysis: Using single case design elements to aid decision making. *School Psychology Quarterly, 14,* 163-181.

Martens, B. K., Hiralall, A. S., & Bradley, T. A. (1997). Improving student behavior through goal setting and feedback. *School Psychology Quarterly, 12,* 33-41.

Martens, B. K., & Kelly, S. Q. (1993). A behavior analysis of effective teaching. *School Psychology Quarterly, 8,* 10-26.

McBride, B. & Schwartz, I. S. (2003). Effects of teaching early interventionists to use discrete trials during ongoing classroom activities. *Topics in Early Childhood Special Education, 23,* 5-17.

McConnell, S. R. (2000). Assessment in early intervention and early childhood special education: Building on the past to project into our future. *Topics in Early Childhood*

Special Education, 20, 43-48.

McGee, G. G., Krantz, P. J., & McClannahan, L. E. (1985). The facilitative effects of incidental teaching on preposition use by autistic children. *Journal of Applied Behavior Analysis, 18,* 17-31.

McNeill, J. H. & Fowler, S. (1999). Let's talk: Encouraging mother-child conversations during story reading. *Journal of Early Intervention, 22,* 51-69.

McWilliam, R. A., Trivette, C. M., & Dunst, C. J. (1985). Behavior engagement as a measure of the efficacy of early intervention. *Analysis and Intervention in Developmental Disabilities, 5,* 59-71.

McWilliam, R. A., Tocci, L., & Harbin, G. L. (1998), Family-centered services: Service providers' discourse and behavior. *Topics in Early Childhood Special Education, 18,* 206-221.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194-216.

Mesibov, G. B., Browder, D. M., & Kirkland, C. (2002). Using individualized schedules as a component of positive behavioral support for students with developmental disabilities. *Journal of Positive Behavioral Interventions, 4,* 73-79.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning of consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). New York: Wiley.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological

assessment: A review of evidence and issues. *American Psychologist, 56,* 128-165.

Neisworth, J. T., & Bagnato, S. J. (1992). The case against intelligence testing in early intervention. *Topics in Early Childhood Special Education, 12*, 1-20.

Nichols, A. & Barnett, D. W. (2004). *Procedures for coding instructional and caring contacts*. Unpublished manuscript, University of Cincinnati.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33,* 271-283.

O'Neill, R. E., Horner, R. H., Albin, R. W., Sprague, J. R., Storey, K., & Newton, J. S. (1997). *Functional assessment and program development for problem behavior: A practical handbook* (2nd ed.). Pacific Grove, CA: Brookes/Cole.

O'Shaughnessy, T. E., Lane, K. L., Gresham, F. M., & Beebe-Frankenberger, M. E. (2003). Children placed at risk for learning and behavioral difficulties: Implementing a school-wide system of early identification and intervention. *Remedial and Special Education, 24,* 27-35.

Parsonson, B. S. & Baer, D. M. (1986). The graphic analysis of data. In A. Poling and W. R. Fuqua (Eds.), *Research methods in applied behavior analysis* (pp. 157-186). New York: Plenum.

Patterson, G. R. (1986). Performance models for antisocial boys. *American Psychologist, 41*, 432-444.

Reinhartsen, D. B., Garfinkle, A. N., & Wolery, M. (2002). Engagement with toys in two-year old children with autism: Teacher selection versus child choice. *Research & Practice for Persons with Severe Disabilities, 27,* 175-187.

Reschly, D. (1988). Special education reform: School psychology revolution. *School Psychology Review, 17,* 459-475.

Reschly, D. J., & Tilly, W. D. (1999). Reform trends and system design alternatives. In D. J. Reschly, W. D. Tilly, & J. P. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 15-37). Longmont, CO: Sopris West.

Reschly, D. J., & Ysseldyke, J. E. (2002). Paradigm shift: The past is not the future. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 3 - 20). Bethesda, MD: NASP.

Risley, T. R., & Cataldo, M. F. Evaluation of planned activities: The PLA-check measure of classroom participation. Lawrence, KS: Center for Applied Behavior Analysis, 1974.

Roberts, J. R., Bailey, D. B., & Nychka, H. B. (1991). Teachers' use of strategies to facilitate the communication of preschool children with disabilities. *Journal of Early Intervention, 15,* 358-376.

Rule, S., Losardo, A., Dinnebeil, L., Kaiser, A., & Rowland, C. (1998). Translating research on naturalistic instruction into practice. *Journal of Early Intervention, 21,* 283-293.

Schwartz, L. S., Carta, J. J., & Grant, S. (1996). Examining the use of recommended language intervention practices in early childhood special education classrooms. *Topics in Early Childhood Special Education, 16,* 251-272.

Shernoff, E. S., Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based

interventions in school psychology: An illustration of task force coding criteria using single-participant research design. *School Psychology Quarterly, 17*, 390-422.

Shinn, M. R. (Ed.). (1998). *Advanced applications of curriculum-based measurement.* New York: Guilford.

Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. I, pp. 671-697). Bethesda, MD: National Association of School Psychologists.

Shinn, M. R., Walker, H. M., & Stoner, G. (Eds.). (2002). *Interventions for academic and behavior problems II: Preventative and remedial approaches*. Bethesda, MD: National Association of School Psychologists.

Smith, L. D., Best, L. A., Stubbs, D. A., Archibald, A. B., & Roberson-Nay, R. (2002). Constructing knowledge: The role of graphs and tables in hard and soft psychology. *American Psychologist, 57*, 749-761.

Smith, M. W. & Dickson, D. K. (2002). *User's guide to the early language & literacy classroom observation toolkit.* Baltimore, MD: Brookes.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. Reading *Research Quarterly, 21*,360-406.

Sterling-Turner, H. E., Watson, T. S., & Moore, J. W. (2002). The effects of direct training and treatment integrity on treatment outcomes in school consultation. *School Psychology Quarterly, 17, 47-77.*

Strain, P. S. & Joseph, G. E. (2004). A not so good job with "Good Job": A response to Kohn

2001. *Journal of Positive Behavioral Interventions, 6,* 55-59.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R., III, Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavioral support and functional behavioral assessment in the schools. *Journal of Positive Behavioral Interventions, 2,* 131-143.

Sugai, G., Horner, R. H., & Gresham, F. M. (2002). Behaviorally effective school environments. In M. R. Shinn, H. M. Walker & G. Stoner, (Eds.). *Interventions for academic and behavior problems II: Preventative and remedial approaches* (pp. 315-350). Bethesda, MD: National Association of School Psychologists.

Thomas, E. J., Bastien, J., Stuebe, D. R., Bronson, D. E, & Yaffe, J. (1987). Assessing procedural descriptiveness: Rationale and illustrative study. *Behavioral Assessment, 9,* 43-56.

Tilly, W. D., Reschly, D. J., & Grimes, J. (1999). Disability determination in problem-solving systems: Conceptual foundations and critical components. In D. J. Reschly, W. D. Tilly, & J. P. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 285-321). Longmont, CO: Sopris West.

Tirapelle, L. & Cipani, E. (1992). Developing functional requesting: Acquisition, durability, generalization of effects. *Exceptional Children, 58,* 260-269.

Torgeson, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40,* 7-26.

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate

and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34,* 33-58.

Turnbull, A., Edmonson, H., Griggs, P., Wickham, D., Sailor, W., Freeman, R., Guess, D., Lassen, S., McCart, A., Park, J, Riffel, L. Turnbull, R., & Warren, J. (2002). A blueprint for schoolwide Positive Behavior Support: Implementation of three components. *Exceptional Children, 68,* 377-402.

Turnbull, H. R., III, Wilcox, B. L., Stowe, M., Raper, C., & Hedges, L. P. (2000). Public policy foundations for positive behavioral interventions, strategies, and supports. *Journal of Positive Behavioral Interventions, 2,* 218-230.

Principles for Reauthorizing the Individuals with Disabilities Education Act. (2003). Retrieved May 12, 2003, from http://www.ed.gov/PressReleases/02-2003/02252003.html

U.S. Department of Education Office of Special Education and Rehabilitative Services (OSERS). (2002). *A new era: Revitalizing special education for children and their families.* Washington, DC.

U.S. Department of Health and Human Services, Administration on Children. Youth and Families/Head Start Bureau. (2000). *The Head Start path to positive child outcomes.* Washington, DC.

VanDerHeyden, A. M., Witt, J. C., & Gatti, S. (2001). Descriptive assessment method to reduce overall disruptive behavior in a preschool classroom. *School Psychology Review, 30,* 548-567.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development of a process for screening

referrals to special education. *School Psychology Review, 32,* 204-227.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391-409.

Vellutino, F.R., Scanlon, D., & Lyon, G.R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against IQ-achievement discrepancy of reading disability. *Journal of Learning Disabilities, 33,* 223-238.

Vellutino, F.R., Scanlon, D., & Tanzman, M. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology, 36,* 367-397.

Vellutino, F.R., Scanlon, D., Sipay, E., Small, S., Pratt, A., Chen, R., & Denkla, M. (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601-638.

Wainer, H. & Braun, H. I. (Eds.) (1988). *Test validity.* Hillsdale, NJ: Erlbaum.

Webster-Stratton, C. (1999). *How to promote children's social and emotional competence.* Thousand Oaks, CA: Sage.

Webster-Stratton, C., Reid, M. J., Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start. *Journal of Clinical Child Psychology, 30,* 238-302.

Witt, J. C., Daly, E., & Noell, G. (2000). *Functional assessments.* Longmont, CO: Sopris West.

Wolery, M. (2000). Recommended practices in child-focused interventions. In S. Sandall, M. E. McLean, & B. J. Smith (Eds.), *DEC recommended practices in early intervention/early*

childhood special education (pp. 29-37). Longmont, CO: Sopris West.

Wolery, M., Anthony, L., Caldwell, N. K., Snyder, E. D., & Morgante, J. D., (2002). Embedding and distributing constant time delay in circle time and transitions. *Topics in Early Childhood Special Education, 22,* 14-25.

Wolery, M., Bailey, D. B., Jr., & Sugai, G. M. (1988). *Effective teaching: Principles an procedures of applied behavior analysis with exceptional students*. Boston: Allyn & Bacon.

Wolery, M., Brashers, M. S., & Neitzel, J. C. (2002). Ecological congruence assessment for classroom activities and routines: Identifying goals and intervention practices in childcare. *Topics in Early Childhood Special Education, 22,* 131-142.

Wolery, M., & Dunlap, G. (2001). Reporting on studies using single-subject experimental methods. *Journal of Early Intervention, 24,* 85-89.

Wolf, T. H. (1973). *Alfred Binet.* Chicago: University of Chicago Press.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49,* 156-167.

Ysseldyke, J. E., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983). Generalizations from five years of research on assessment and decision making: Findings of the Minnesota Institute. *Exceptional Educational Quarterly, 4,* 75-93.

Table 1

Summary of Classes of Technical Checks for 3 Tiers

Tier 1: *Curriculum: Validity of scientifically-based curriculum; fidelity of implementation of curriculum.* *Screening and assessment*: Validity and reliability of universal screening methods; validity and reliability of progress monitoring. *Instructional environment*: Validity of instructional and managerial methods; fidelity of instructional and managerial implementation. *Problem solving*: Fidelity of problem solving for teachers and support personnel; procedural fidelity for decision rules; validity and reliability of decision criteria for moving to Tier 2.
Tier 2: *Embedded and small group interventions.* Validity of interventions; fidelity of implementation; progress monitoring through interventions linked to Tier 1 objectives; decision rules for increasing/decreasing intensity within and between tiers (same procedures as Tier 1); social validity of goals, methods, and outcomes.
Tier 3: *Intensified and individualized*. Technical adequacy of problem solving for individual child to include validity and reliability of target variable selection and occurrence, intervention selection, intervention adherence, and outcome determination; decision rules for increasing/decreasing intensity. Social validity of goals, methods, and outcomes.

Table 1(con't)

Summary of Classes of Technical Checks for 3 Tiers

Tier 3: *Eligibility for specialized services.* Intensity estimates are based on a data set meeting criteria: specifying and scheduling interventions (i.e., duration of contacts between change agent and child) and examining level and trend of progress across 3 tiers of structured intervention trials with data demonstrating that special services would be necessary for further progress due to the degree to which interventions differ from typical routines in terms of resources, time, involvement of professionals beyond the child's teacher, and other factors.

Table

Examples of 3 Tiers of Interventions at Head Start

Tier 1: Examples of class-wide Interventions

Curricular interventions: adoption of a strong, sequenced, early literacy and language curriculum; adoption of specialized curriculum (e.g., for social behaviors or language); new activities (environmental enrichment) and activity management (introducing, modeling, and rotating activities).

Feedback to teachers: on key instructional variables such as children's engagement, positive scanning and vicarious reinforcement, results of standardized classroom assessments of activities and instructional techniques.

Intentional and incidental instruction: increasing children's opportunities to respond; interspersing hard and easy material, interactive reading, choral responding, modeling, wait time, pacing; incidental teaching for social and language skills.

Classwide management: entry routines, effective transitions, limits or guides for activities, consequence hierarchy; choice making.

Tier 2: Embedded and Small Group

Language: Milieu language interventions embedded in routines to increase practice opportunities such as missing item format; brief natural time delay

Social: Curriculum as a pull out program for groups to learn and practice skills appropriate for classroom practice.

Table 2 (con't): Examples of 3 Tiers of Interventions at Head Start

---

Tier 3: Intensified and Individual

Structures for instructional and/or behavioral interventions: Individualized (activity) schedules; peer or buddy systems; social stories; choice making; planned setting events. *Idiosyncratic empirical interventions* (based on functional assessment/ hypotheses or reinforcement preference assessment)*:* contingent reinforcement, non-contingent reinforcement (fixed time schedules); high probability request sequences, neutralizing routines; consequence hierarchy; intensified language (increasing intervention occasions by adding embedded milieu language interventions).

Figure Captions

Figure 1. Examples from permanent products checklist.

Figure 2. Graph showing tiered progression from class, to group, to individual data.

Tier 1: School- and Class-wide

Date	*Activity*
	Communication plan with teacher support team.
	Plan for measuring instructional environment.
	Agency universal screening on Curriculum Based Measurement (CBM; based on the Head Start Outcomes Framework), Fall, Winter & Spring.
	Graph showing progress monitoring by classroom for children scoring below 25%tile using universal screening results as benchmarks, and decision rule use.
	Teacher receives data-based feedback about instruction and instructional environment, child engagement, and CBM results with benchmark comparisons; goals are set.
	Plan for teacher support as necessary to possible include class-wide instruction, curricular interventions, management, routines, or other supportive interventions (staff development, smaller group, and individualized teacher support); determination of required resources, to accomplish progress monitoring for instructional objectives & screened students.
	Summary of effects of scripted interventions as a graph; description of decision rules to move children to Tier 2.

Tier 2: Embedded and Small Group

	Select target students whose progress under Tier I interventions is not sufficient to reach year end goals by *dual criteria decision rules* data (level, trend) and graphed data.
	Identify targeted instructional activities and child/instructional variables (i.e., embedded or small group).
	Select instructional/other variables for progress monitoring as needed; 1st Tier data serves as baseline.
	Team reviews data and makes further plans (increasing/Tier 3 or decreasing/fading intervention intensity (Tier 1); intervention maintenance or fading.

Tier 3: Intensified and individualized

	Team is comprised of parents, appropriate professionals from preschool agency and child's local school district.
	Complete a record review.
	Select instructional and contextual variables for RTI measurement (Barnett, Bell, et al., 1999).
	2nd Tier data serves as baseline; develop & carry out additional baseline observations as needed for selected variables and functional discrepancies.
	Form functional hypotheses related to individual data; set goals; develop script for interventions; select increasing or decreasing intensity design and rationale; refine or revise plans & provide support as necessary; plan for progress monitoring; plan technical checks for targeted variables and interventions (e.g., co-observations of child measures & intervention adherence).
	Add specific coaching phase from teacher support team as needed.

	State decision rules with data for moving to Tier 1 or 2, or remaining in Tier 3; for child remaining in Tier 3, review Federal entitlements & state rules; review cumulative data on functional discrepancy and intervention intensity for special services eligibility*;* state decision criteria for eligibility for specialized services as a child with a disability; data across 3 tiers is used for eligibility; recent data is used to plan IEP with objectives from Tier 1.

Notes: This figure illustrates items from a comprehensive procedural checklist. A complete checklist available on request from 1st author.