

Running Head: ELIGIBILITY DETERMINATION FOR CHILDREN WITH SOCIAL
EMOTIONAL AND BEHAVIOR PROBLEMS

Determination of early intervention eligibility based upon deficits in acquisition of social
and self-regulatory behavior: research foundation for a developmental model of early
behavior problems

Hsiang-Yeh Ho

Stephen J. Bagnato

University of Pittsburgh / Children's Hospital of Pittsburgh

Ho, H., & Bagnato, S.J. (2007). Determination of early intervention eligibility based upon deficits in acquisition of social and self-regulatory behavior: research foundation for a developmental model of early behavior problems. Pittsburgh, PA: TRACE Center for Excellence in Early Childhood Assessment, Early Childhood Partnerships, Children's Hospital/University of Pittsburgh; US Department of Education, Office of Special Education Programs, and Orelena Hawks Puckett Institute.

Abstract

Four types of practices were identified for eligibility determination based on deficits in acquisition of social emotional, behavior, and self-regulatory problems: screening assessments, curriculum-based assessments, diagnostic assessments, and response to intervention models. Ten practice characteristics were identified to examine the technical adequacy, classification accuracy, and usability for determining eligibility. The practice characteristics included: consensus decision making (professional and professional-parent consensus); functional assessment (observation, authentic ecological context, multiple times or settings); administration reliability (training, inter-rater reliability); pre-referral intervention; linking; and utility. A total of 16 studies were selected. The findings suggest that more studies are needed to support the efficacy and usability of the identified practices in determining eligibility, especially with younger children. Implications for practice and research are suggested.

Purpose

“One can no more classify children into discrete categories than one can point to the exact spot in a sunset where red changes to orange (Braden, 1989).”

Braden (1989) clearly expressed the difficulties and challenges in making eligibility decisions for young children, especially those with social and self-regulatory problems. Unlike children with physical disabilities (i.e., deaf, blind), it is difficult to pinpoint at which degree a child’s aggressive behavior or level of emotional disturbance is serious enough to require intervention. According to federal policy, children with social emotional, behavior or self-regulatory problems are able to access to Part C (birth to three) services by: 1) showing measurable developmental delay in the social emotional domains, 2) having a diagnosed physical or mental condition which has a high probability of resulting in a developmental delay or; 3) classifying as at risk, at the state’s discretion (Individual with Disabilities Education Act Amendments, 1997). For Part B (3 to 5) eligibility criteria, children have to be identified with serious emotional disturbance (SED/ED), or show developmental delay, at the state’s discretion, in the social emotional domain (IDEA, 1997). However, when developmental delay (DD) is used for identification purposes, children with social emotional and behavioral problems are often overlooked for several reasons.

First, many parents and professionals believe that early social emotional and behavior problems are transient, which may become a barrier that eliminates children’s chances to receive intervention services as early as possible. For example, children with Emotional or Behavioral Disorders (EBD) were usually not identified until their late elementary school years (Kaufmann & Perry, 2001; Landrum, 2000). Second, when assessments are driven by eligibility policy, children with social emotional and behavior problems are easily over-looked (Evangelista &

McLellan, 2004; Webster-Stratton & Hammond, 1997). For example, prevalence studies suggested that 12 to 16% of one to two year-olds show a significant delay in social emotional development (Briggs-Gowan, Carter, Skuban, & Horwitz, 2001; Conroy & Brown, 2004), and 37% of those children continue to exhibit problem behaviors into preschool years. In the school-aged population, the prevalence rate was found to be 3 to 6% but only 0.5 to 1% of those children were identified. Studies also suggested that early emerging social emotional and behavioral problems seem to have long-term effects for children in terms of development and achievement (Campbell, 1991). Third, the construct of social emotional and behavior problems have no objective reality and are usually defined within professional contexts (Kauffman, 2005; Kaufmann & Perry, 2001). However, to meet the eligibility criteria, children's behaviors have to be quantified by using a standardized, norm-referenced measurement (i.e., 1- or 1.5 standard deviation above or below the mean) to indicate the level of deficits that a child exhibits.

The application of evidence-based practices for early identification and eligibility determination of young children with social emotional and behavior difficulties is of great interest to researchers, practitioners, and policy makers. This article represents a synthesis of the state of knowledge and evidence concerning the early intervention eligibility determination for children who have social emotional, behavior, and self-regulatory problems. The purpose of this synthesis is twofold, to identify: 1) practices that are used for determining eligibility for young children with social and self-regulatory problems, including assessments (i.e., screening, diagnostic, curriculum-based) and problem-solving based models, and 2) the characteristics of measurements and models that are useful for eligibility decisions. The studies reviewed were examined based on the practices that were associated with the accuracy of the identification, the usage and reliability of the assessment tools, and the utility of the study outcomes.

Background

Description of the Practices

Three methods are commonly found in the literature pertaining to early intervention eligibility. The first method is a “measurement oriented” approach, which means that eligibility decision relies heavily on the results of one or more measurements. Such a method usually involves two types of assessments from identification to decision making (1) Screening assessment, used for early identification of targeted children; (2) Diagnostic assessment, an in-depth assessment (such as a “curriculum-based” approach) which links assessment to the planning and implementation of intervention strategies; and (3) a “problem-solving model” that is systematically applied in school settings within a multi-tiered structure, in which the intervention is provided before children are identified for services. These practices are described in the following sections.

Screening assessments

Screening is defined as a brief assessment procedure designed to identify children who should receive more intensive diagnosis (Meisels, 1988). Researchers recommend that the developmental screening instruments should be brief, norm-referenced, inexpensive, standardized in administration, objectively scored, broadly focused across all areas of development, reliable and valid (Kauffman, 2005). Screening should be as accurate as possible so that the negative consequence of over-referral and under-detection are minimized (Glascoe, 2005). A review of the relevant literature suggests that an effective measure should (1) be accurate which is reflected by the sensitivity, specificity, and positive predictive value (Glascoe, 2005); (2) have multiple gates so that children who are at risk (i.e., environmental, medical,

familial) can be monitored, and; (3) include multiple individuals across multiple times, and engage family members in the decision making process (Hill, Coie, Lochman, & Greenberg, 2004). Although screening tools are not directly linked to eligibility decisions, they play a crucial role in early identification of children in need of early intervention.

Diagnostic Assessments

Eligibility decisions are most often based on the findings of diagnostic instruments. Without a doubt, the quality of the instrument is crucial in accurately identifying children. The Zero to Three Diagnostic Classification System (DC: 0-3) is one example of a frequently used and developmentally sound diagnostic instrument. It includes both assessment procedures and a diagnostic classification based on evaluation findings (Dunst, Storck, & Snyder, 2006). The system is a landmark development in the design of a diagnostic framework for infant and early childhood behavior disorders that is developmentally-appropriate, intervention-based, and compatible with best practices in early intervention (Bagnato & Neisworth, 1999). The DC: 0-3 applies a systematic approach to identifying developmental and mental health problems in children (Bricker, Davis, & Squires, 2004), and underscores the view that early social emotional and behavioral problems cannot be classified as traditional mental health diagnoses until the effect of intervention or treatment are documented (Bagnato & Neisworth, 1999).

There are five axes in the DC: 0-3. Axis I, *Primary Diagnosis*, is related to the diagnosis of the pathology affecting the child. The intent of the DC:0-3 is to describe behaviors rather than symptoms, which makes it more suitable for infants and toddlers (Guedeney et al., 2003). For example, Regulatory Disorders and Multi-System Developmental Disorders are two new categories which are highly relevant for infants and toddlers. Axis II, the *Relationship Disorder Classification*, enables the classification of interactive disorders between parents and children.

Axis III, *Medical and Developmental Disorders and Conditions*, is exclusively medical and developmental. Axis IV, *Psychosocial Stressors*, assesses the impact of vital events in child's life in relation to the protective potential provided by the environment, mainly within the scope of the parent-child relationship (Guedeney et al., 2003; Minde & Tidmarsh, 1997). Axis V, *Functional Emotional Developmental Level*, gives a picture of the functioning of the infant in relation to his/her acquisition of the ability to express and regulate emotions, appropriate to their age group (Guedeney et al., 2003). By using these five axes, assessors will be able to identify risk and resilience actors that contribute to designing the therapeutic intervention. In such scenarios, the assessment is linked to the intervention.

Curriculum-Based Measurements

Curriculum-based measurements (CBM) are defined as “a form of criterion-referenced measurement wherein curricular objectives act as the criteria for the identification of instructional targets and for the assessment of status and progress” (Pretti-Frontczak, 2002). CBMs are designed to strengthen the link between assessment and intervention to improve outcomes for young children and families. CBMs are composed of meaningful assessment items, permit a range of performance criteria that represent a child's functional behavioral repertoire, allow for modifications and adaptations, provide procedures for family involvement and facilitate collaboration among team members (Bagnato & Neisworth, 1999; Bricker et al., 2004). Bagnato and colleagues (1997) summarized linking standards that CBMs should be judged on (a) authenticity, (b) convergence, (c) collaboration, (d) equity, (e) sensitivity, (f) congruence, (g) utility, and (h) acceptability (Bagnato, Neisworth, & Munson, 1989).

Previously, CBMs were not widely used for eligibility determination because they did not yield norm-referenced scores. Macy and Hoyt-Gonzales (2007) argue that the characteristics of

CBMs are actually beneficial for determining eligibility. First, items from CBMs are functional and can be easily translated into programmatic efforts. Functional skills are easily mapped onto daily routines and activities, so when a CBM is administered as the eligibility assessment, meaningful authentic interventions can be generated. Second, a CBM provides baseline of information that can be used to evaluate individualized programs over time. A CBM allows for the comparison of one's performance against him/herself rather than comparisons made to a peer group. The information collected is not only useful for determining eligibility but also serves this comparison function (Macy & Hoyt-Gonzales, 2007).

Response to Intervention and Response and Recognition

Response to Intervention (RTI) was initially intended for use in determining the IDEA eligibility category of Specific Learning Disability. The improvements in IDEA associated with the use of the RTI eligibility determination model are also being applied to evaluations for other eligibility categories including Emotional Disturbance (Fairbanks, Sugai, Guardino, & Lathrop, 2007; Gresham & REACH Project, 2005; Harris-Murri, King, & Rostenberg, 2006). Interested readers are referred to Coleman, et al. (2007) to study the efficacy of RTI models. On the basis of what is known about Response to intervention (RTI) for school-aged children, an early intervening system called Recognition and Response (R&R) is being developed and validated for use with younger children (i.e. 3-5 year-olds) in various types of early education settings. Recognition refers to the methods used to identify young children who exhibit early learning difficulties and who may be at risk for learning disabilities later on. Response refers both to the ways in which teachers and parents respond to children with learning difficulties as well as to the ways in which young children react to specific interventions (Coleman, Buysse, & Jennifer Neitzel, 2006). Even though the initial purpose of R&R and RTI was to identify children with

learning disabilities, researchers have suggested application of this model to identification of children with social emotional and behavioral problems.

The R&R model is built on the earlier work of RTI in important ways by (1) focusing on helping teachers support children's academic as well as their social-emotional development; and (2) incorporating other essential elements from the RTI model into an early intervention system (i.e., linking screening, assessment, and intervention together, monitoring progress with evidence-based curriculum, and collaborative decision making) (Coleman et al., 2006). The R&R system emphasizes a systematic approach to respond to early learning difficulties that includes assessing the overall quality of early learning experiences for all children. As a result, it does not rely on formal diagnosis and labeling. The R&R model is a decision-making model in which services are delivered and altered based on children's reaction to the intervention (Barnett et al., 2006; Harris-Murri et al., 2006). The R&R model includes four essential components (1) Intervention hierarchy; (2) Screening, assessment, and progress monitoring tiers; (3) Research-based curriculum, instruction, and focused interventions; and (4) Collaborative problem-solving process for decision making. These components are based on the premise that all students receive research-based instruction in the general education setting, and that student learning is assessed early and often. Eligibility for EI/ECSE (early intervention, early childhood special education) is determined only when children show "unresponsiveness" to general education instruction (Harris-Murri et al., 2006).

Practice characteristics

The selected studies were appraised using six practice characteristics that were considered useful for (1) identifying children who are at risk for or with delay in social-emotional,

behavioral, and self-regulatory domains, (2) documenting the accuracy and reliability of a measurement or model for determining targeted children, and (3) effectively using measurements or models to assess children's skills and progress. The six evidenced-based practice characteristics were: (a) *consensus decision making*, including professional collaboration and professional-parent collaboration, (b) the use of *functional assessment* in the procedure, including ecological observation contexts across multiple settings and times, (c) *administration reliability*, the construction of inter-rater reliability between raters and training for administrators in the methods, (d) the administration of the *pre-intervention support*, (e) *linking* between assessment and intervention, (f) *utility* in terms of the research outcomes. In the following sections, the term *targeted behaviors* was used to refer to children's social emotional, behavioral and self-regulatory problems.

Term definition

Consensus decision making was recognized when there was more than one individual working collaboratively in the decision making with regard to children's developmental status, progress, or needs. Individuals may be professionals, researchers, teachers, and parents. Professional consensus was identified when the decision was made by more than one professional or an interdisciplinary team to make the eligibility decisions. Professional-parent consensus was recognized when at least one main caregiver and one professional were involved in the decision making.

Functional assessment refers to how a child actually functions in the natural context. It involved three criteria: (1) observational data was considered as part of assessment, (2) observations were conducted in an authentic ecological context, such as home, classroom, or play ground, (3)

observation or assessment was conducted across multiple settings or multiple times, in order to monitor progress.

Administration reliability was crucial when there was more than one individual involved in the administration. Two characteristics were identified (1) inter-rater reliability was conducted, and (2) training was provided for individuals or raters (i.e., administering tools or observations) to ensure that individuals have the same degree of knowledge and skills when conducting the study.

Pre-referral intervention was recognized when the intervention was provided before determining the eligibility status. This characteristic also provided information regarding children's progress over time.

Linking refers to the linkage between assessment and intervention; the results of the assessment was used for planning intervention strategies. This characteristic is especially useful and crucial for Part C early intervention programs when considering cost effectiveness.

Utility was recognized when the study or assessment outcomes were useful for the eligibility determination or for intervention strategies, such as the reports of identification accuracy, technical adequacy, or the effectiveness of a measurement or a model.

Search Strategy

Search terms

In order to capture all the possible social emotional, behavior, and self-regulatory behaviors and problems identified in the early childhood, multiple terms were used including common early childhood diagnoses (i.e., *ADHD*, *autism*, *Oppositional Defiant Disorder*) and universal

terms (i.e., *social emotional/behavior problems, social/emotional disturbance, self-regulatory problems, developmental delay*). The following terms: *eligibility, early intervention, early identification, developmental delay, response to intervention, recognition and response, developmental screening, systematic screening, curriculum-based assessment, diagnostic assessment, and intervention-based assessment* were used and combined interchangeably to identify studies. These terms were used to identify relevant published and unpublished studies, including position papers, literature reviews, and empirical studies.

Search Sources

Internet literature searches were conducted and the published literature was surveyed through databases from the fields of psychology, education, special education, developmental disabilities, child development, and early intervention. The primary databases searched were: Psychological Abstracts online (PsychoINFO), Education Resource Information Center (ERIC), Ovid, MEDLINE, Web of Science and Google search. In addition, hand searches were done of the reference sections of relevant studies, review articles, book chapters, and books. Repeated sweeps of various electronic databases reference sections of newly identified studies were examined until no further investigations were located.

Selection criteria

In the following section, several terms were identified based on the interests of this synthesis. First, the social emotional, behavior, and self-regulatory behaviors were identified as *targeted behaviors*, and children who were under 5 years old and demonstrated targeted behaviors were referred as *targeted children*. The *identified practices* were the practices that have been discussed

in the previous section. When reviewing the literature, no study was found to use any of the identified practices to examine the eligibility determination decisions on children with targeted behaviors for Part C early intervention. The selection criteria was extended and the studies were included when they: (1) examined the classification function of the identified practices with targeted children, and the results showed good to excellent classification rate; (2) investigated the technical adequacy of the identified practices in targeted children, and the results showed moderate to high levels of adequacy; and (3) explored eligibility determination by using the identified practices, and targeted behaviors were explored.

Search Results

The total number of subjects who participated in the 16 studies was 9,953. Four types of sample groups were identified: (1) community samples, consisted of 7,963 children (80%), (2) school samples, comprised 998 children (10%); (3) clinical samples, comprised 766 children (approximately (8%), and (4) preschool samples, comprised 226 children (2%). Table 1 shows the demographics of the studies. The age of the children ranged from birth through eight years old. Thirteen of the studies reported gender; boys made up 42 to 80% of the sample. Fourteen studies reported Caucasians as the majority ethnic group (45 to 86%) with the exception of one study in which the participants were all of Hispanic ethnicity. Seven studies (43.8%) reported families' socioeconomic status (SES). Three of these studies mainly recruited children of low SES, and four studies included children from diverse backgrounds.

Type of study

Dividing all 16 studies into groups based on the type of practices, there were six Screening Assessments (SAs), two Curriculum-based Assessments (CBAs), four Diagnostic Assessments (DAs), and four Response to Intervention Models. These studies were divided into four categories for the focus of the studies: Early identification (EI, $n = 2$), classification accuracy (CA, $n = 4$), technical adequacy (TA, $n = 6$), and pre-referral intervention (PI, $n = 4$). The first set of studies focused on examining the efficacy of using a tier-structured screening process for early identification. The second set of studies examined the accuracy of the classification function of measures. The third set investigated the technical adequacy in terms of reliability and validity of the measures, or diagnostic system. Finally, the last set consisted of studies that introduced interventions in a tier-structured strategy, and the results of the intervention were connected to eligibility decisions. Table 2 shows the detailed categorization of practice types.

Research Design

Three types of research design were generalized: (1) Correlational design ($n = 10$), (2) Quasi-Experimental design ($n = 2$), and (3) Quasi-Experimental Longitudinal design ($n = 4$). The quasi-experimental studies consisted of a hierarchical structure in the design, which means students moved to the next stage depending on how they performed in the first stage. The length of time between the initial to the last stage ranged from 6 to 24 months.

Study Outcomes

Study outcomes were examined in terms of (1) classification accuracy (i.e., specificity, sensitivity), (2) the technical adequacy (i.e., reliability, validity), and (3) the usability of the practice in determining eligibility. According to Cicochetti (1995), the sensitivity or specificity

was identified as (1) “poor”, when it was below 70%; (2) “fair”, when it was between 70% to 79%; (3) “good”, when it was between 80% to 89%; and (4) excellent, when it was equal or above 90% (Cicchetti, 2001).

Synthesis Findings

Practice Characteristics

Professional Consensus

Consensus decision making entails the integration of information from different persons (i.e., parents, teachers, professionals) or resources to facilitate appropriate decision making in terms of diagnosis or intervention planning. Overall, 25% (n = 4) of studies included *Professional Consensus* while only 6.3% (n = 1) included *Professional-parent Consensus* in either decision making or when determining children’s developmental status. None of the SA or CBM studies included parents in the assessment process. One of the four DA studies included both the *Professional-* and *Professional-parent consensus* in decision making; while 3 of the four RTI studies included *the Professional consensus* but not *Professional-Parent consensus*. Professionals included in the studies were researchers, psychologists, postdoctoral psychologists, psychiatrists, social workers, and clinicians.

Functional Assessment

Functional Assessment facilitates the understanding of a child’s real functioning by observing the child in an ecological context in multiple settings across multiple times. Through functional assessment, investigators are able to better understand a child’s strengths, weakness,

the frequency and intensity of the behaviors etc., which are crucial when planning the intervention. Of 16 studies, 81.3% (n = 13) included this characteristic in the evaluation process. More specifically, across different types of practices, 50% to 100% of the studies included observation, 75% to 100% of the studies conducted assessments in an authentic context, and 50% to 100% of the studies conducted assessments across multiple times or settings.

Administration Reliability

This characteristic included two components: *Training* and *Inter-rater reliability*, which ensured the administration of the assessment, was reliable and consistent. Seventy-five percent (n = 12) provided training for administrators and 50% (n = 8) conducted inter-rater reliability (ranging from 80 to 90%, between raters). In each type of assessment, 50% of the SAs provided training while 66.7% of SAs conducted inter-rater reliability when there were more than two raters administering the assessment. Training was provided in all of the studies which used CBM and DA, but the percentage of those measuring inter-rater reliability was 25% in the DAs and 50% in the CBMs. In the RTIs, the training was provided in 75% of the studies, while only 50% conducted inter-rater reliability.

Pre-referral Intervention

Out of the 16 studies 31.3% (n = 5) provided intervention. None of the studies which used CBM also utilized pre-referral intervention in the process of assessment. Only 25% of DA studies included this characteristic; while 100% of RTI studies did so.

Linking

Thirty-six percent ($n = 6$) of the studies applied this practice characteristic. It was observed that none of the SAs and CBMs included this characteristic, while 75% in DAs to 100% in RTIs included the linking function.

Utility

Among 16 studies, 56.3% ($n = 9$) of the studies indicated the practices were useful for determining eligibility based on their findings. None of SAs informed utility, while 75% to 100% of CBMs, DAs, and RTIs included this characteristic. Table 3 shows the distribution of the identified characteristics in the selected studies, and Figure 1 provides a graph of the distribution.

Effectiveness of the Practices

Based on the findings in this synthesis, the effectiveness of each practice is discussed separately in the following section. Table 4 shows the major findings and results of the selected studies.

Effectiveness of the Screening Assessments

Early identification was the main purpose of using SAs. SAs alone were not able to inform eligibility; however, they were influential in the initial stage of the assessment process, especially for at-risk populations. The effectiveness of SAs was examined in terms of the predictive accuracy and technical adequacy. Five of the six studies reported predictive accuracy of the assessments. Results showed that the range of the sensitivity was from 63 to 100% ($M = 85\%$), and the range of the specificity was from 63 to 97% ($M = 89\%$). Considering the purpose of SAs, usually higher specificity is desired. Feil et al. (1993) examined a tier-structured

screening measurement, and the results showed that it had the better classification accuracy comparing to others.

Four of the six studies reported technical adequacy. The average inter-rater reliability ranged from .38 to .97 across different measures ($n = 4$); test-retest reliability ranged from .49 to .94 ($n = 3$), concurrent validity ranged from .25 to .91 ($n = 2$). The results of technical adequacy across different studies seemed to be even. Functional assessment (83.3 to 100%) and administration reliability (50 to 66.7%) were most often included in the studies, while no studies included consensus decision making, pre-referral intervention, linking, or utility. (See table 3.) These findings are reasonable when considering the main purpose of using screening assessments – early identifying targeted children rather than inform decision making.

Effectiveness of the CBM Assessment

Both CBM studies examined the predictive accuracy, technical adequacy, and the usability for informing eligibility decisions. As opposed to screening assessments, CBMs are designed to inform eligibility decisions; therefore, high sensitivity is desired. Findings in CBMs were encouraging. The sensitivity ranged from 88.7 to 100 percent ($M = 93\%$, $n = 2$). The specificity ranged from 75 to 89 percent ($M = 80.6\%$, $n = 2$). Concurrent validity was from .42 to .65 ($n = 1$), and interrater agreement ranged from .80 to .89 ($n = 1$), which showed adequate technical adequacy.

Effectiveness of the Diagnostic Assessment

One of the six studies reported inter-rater reliability (90%), one reported 7 months interval test retest reliability (.51 to .84), and one showed concurrent validity (-.24 to .81). Seventy-five

percent ($n = 3$) of the selected studies examined the characteristics in terms of differences and similarities between DC: 0-3 and DSM-III/R, V. The results found in this synthesis supported the findings in Dunst et al.'s (2006) synthesis article. Even though the DC: 0-3 system was more developmentally appropriate for young children compared to the DSM system, more studies are needed to support its efficacy in eligibility decision making (Dunst et al., 2006).

Epstein et al. (2002) examined the usability of a tier-structured measurement for kindergarteners and first graders. The results showed moderate to high level of concurrent validity. It is evident in Table 3 that nearly all of the practice characteristics are included in the studies; however, consensus decision making and pre-intervention are utilized less often.

Effectiveness of the Response to Intervention Models

One of the four studies reported the sensitivity (33%), specificity (99%), positive predictive value (50%), and the negative predictive power (97%) of the RTI model. It is important to recognize that only one study examined the eligibility decisions on children with social emotional and behavior problems by using RTI. Others were focused on learning disabilities, academic problems, or reading disabilities. The reason to include these studies was to provide information for readers to understand how this newly developed model is implemented in schools, and how it could be utilized within an early childhood system. Findings, in general, were impressive; however, for determining eligibility, higher sensitivity is needed.

When looking at the Table 3, it is clear that each characteristic was equally considered in the studies. Linking, pre-referral intervention, and consensus decision making, which were commonly missing in the previous practices, were all included in the RTIs. In other words, RTIs

implemented and integrated evidenced-based practice characteristics, which is encouraging; however, empirical studies are still needed to support the usability of this model for determining eligibility, especially in the area of social emotional and behavior problems.

Conclusion

There are two main findings in this synthesis. First, there are several well-developed and valid screening measurements that can be used for early identification of very young children with social emotional, behavior, or self-regulatory problems. Second, findings were encouraging in terms of eligibility determination for children with social emotional, behavior and self-regulatory problems (targeted behaviors), such as the development of the CBM assessments (i.e., AEPS: E) (Bricker, Yovanoff, Capt, & Allen, 2003; Macy, Bricker, & Squires, 2005), diagnostic measurements (i.e., SAED, DC: 0-3) (Epstein, Nordness, Cullinan, & Hertzog, 2002; Thomas & Clark, 1998), and RTI (Fairbanks et al., 2007). These measurements appear to be valid, reliable, and accurate for the purpose of classification and identification of eligibiliyt. However, limited by the numbers of studies for each type of practice or measurement, the evidence is not strong enough to support their usability for eligibility determination. The average number of studies under each practice was 3.3, ranging from 2 to 6. Among these studies, 12 studies focused on the deficits of acquisition in targeted behaviors, and 11 studies included children less than five years old. Under each type of practice, the average number of studies investigating the same measure ranged from one to three. For example, two studies investigated the classification function of the ECBI; two examined the AEPS: E; three examined the DC: 0-3; and one examined the SAED. More studies are still needed to investigate the identification and classification function of these measures in terms of eligibility decisions.

Overall, it is clear that there is not enough evidence to support the practices for determining eligibility based on the acquisition of social emotional, behavior, and self-regulatory problems. This finding is not surprising, but alarming when considering children's developmental outcomes, especially for those who are struggling with their poor social skills, emotional disturbance, and behavior problems. Researchers have observed that because of the difficult nature of identifying social emotional and behavioral problems in young children and concerns in labeling, children are may not be identified until their problems become extreme. In addition, evidence exists that many of the students who would benefit from EBD services are actually placed in programs for students with learning disabilities (Kauffman, 2005). This misplacement is partially due to the lack of well-developed instruments.

Implication for Practice

The ten practice characteristics provide a framework for professionals and teams to assess children's strengths and weakness, functional behaviors, and risk factors in a reliable, comprehensive, and meaningful manner. Most studies included functional assessment and administration reliability, however, consensus decision making was a characteristic often missing in the studies, especially professional-parent consensus. According to IDEA, family involvement is important when making eligibility decisions, not only because parents can provide pertinent information about young children's behavior at home, but also because the effectiveness of intervention without family involvement is often limited. The RTI models included most practice characteristics, which suggests its potential usage for eligibility determination.

Finally, there is a trend of developing tier-structured practice when conducting screening and assessments. The studies included in this synthesis show that a multi-gating screening procedure

had excellent, almost 100%, sensitivity and specificity. In these multi-gating screening studies, teacher rankings and nominations were used in the first gate, teacher ratings was conducted at the second stage, and observation by trained professionals was the last stage. Similar procedures were observed in the RTI models, and the results also showed classification accuracy. It is clear that by using a tier-structured approach, regardless of the screening or assessment process, the hierarchical structure increases the sensitivity and specificity as the tier goes higher.

Implication for Research

This synthesis provides evidence for some potential measurements for early identification children with social emotional and behavioral problems and informing eligibility decisions. For example, Macy et al. (2005) and Bricker et al., (2003) examined the usability of AEPS: E for eligibility determination. However, the efficacy of using the AEPS: E for children specifically with social emotional and behavior problems is unknown. In addition, the development of RTI models expresses a clear trend that researchers have shifted their attention from eligibility driven assessment and “wait-to-fail” models to develop more systematic, comprehensive, and intervention-prevention based approaches and practices to assess and monitor children’s developmental status, and emphasize an understanding of children’s strengths, weaknesses, and needs to determine who can benefit from additional prevention and early services. Fairbanks and colleagues (2007) study was the only study to focus on children’s problem behaviors, and it provides a blueprint and direction for future researchers. Directions for future study should focus on constructing the psychometric data for the selected measurements (i.e., DC: 0-3, SAED, AEPS: E, etc.) to provide the evidence of the usability of these measurements in order to identify

and assist children with social and emotional problems and difficulty with self regulation and behavior.

References

- Bagnato, S. J., & Neisworth, J. T. (1999). Collaboration and teamwork in assessment for early intervention. *Child Adolesc Psychiatr Clin N Am*, 8(2), 347-363.
- Bagnato, S. J., Neisworth, J. T., & Munson, S. M. (Eds.). (1989). *Linking Developmental and Early Intervention: Curriculum-based Prescriptions* (2nd ed.). Rockville, Maryland: ASPEN Publication.
- Barnett, D. W., Elliott, N., Wolsing, L., Bunger, C. E., Haski, H., McKissick, C., et al. (2006). Response to intervention for young children with extremely challenging behaviors: What it might look like. *School Psychology Review*, 25(4), 568-582.
- Bricker, D. D., Davis, M. S., & Squires, J. (2004). Mental health screening in young children. *Infants and Young Children*, 17(2), 129-144.
- Bricker, D. D., Yovanoff, P., Capt, B., & Allen, D. (2003). Use of a curriculum-based measure to corroborate eligibility decisions. *Journal of Early Intervention*, 26(1), 20-30.
- Briggs-Gowan, M. J., Carter, A. S., Irwin, J. R., Wachtel, K., & Cicchetti, D. V. (2004). The Brief Infant-Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *J Pediatr Psychol*, 29(2), 143-155.
- Briggs-Gowan, M. J., Carter, A. S., Skuban, E. M., & Horwitz, S. M. (2001). Prevalence of social-emotional and behavioral problems in a community sample of 1- and 2-year-old children. *J Am Acad Child Adolesc Psychiatry*, 40(7), 811-819.
- Campbell, P. H. (1991). Evaluation and assessment in early intervention for infants and toddlers. *Journal of Early Intervention*, 15(1), 36-45.

- Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, 23, 695-700.
- Coleman, M. R., Buysse, V., & Jennifer Neitzel, J. (2006). *Recognition and Response: An early intervening system for young children at-risk for learning disabilities*. Chapel Hill: The University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Conroy, M. A., & Brown, W. H. (2004). Early identification, prevention, and early intervention with young children at risk for emotional or behavioral disorders: Issues, trends, and a call for action. *Behavioral Disorders*, 29(3), 224-236.
- Dunst, C. J., Storck, A., & Snyder, D. (2006). Identification of Infant and Toddler Social-Emotional Disorders Using the DC: 0-3 Diagnostic Classification System. *Cornerstones*, 2(2), 121.
- Epstein, M. H., Nordness, P. D., Cullinan, D., & Hertzog, E. (2002). Scale for assessing emotional disturbance: Long-term test-retest reliability and convergent validity with kindergarten and first-grade student. *Remedial and Special Education*, 23(3), 141-148.
- Evangelista, N., & McLellan, M. J. (2004). The Zero to Three Diagnostic System: A framework for considering emotional and behavioral problems in young children. *School Psychology Review*, 33(1), 159-173.

- Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M. (2007). Response to intervention: Examining classroom behavior support in second grade. *Council for Exceptional Children, 73*(3), 288-310.
- Feil, E. G., & Becker, W. C. (1993). Investigation of a multiple-gated screening system for preschool behavior problems. *Behavioral Disorders, 19*(1), 44-53.
- Feil, E. G., Walker, H., Severson, H., & Ball, A. (2000). Proactive screening for emotional/behavioral concerns in Head Start preschools: Promising practices and challenges in applied research. *Behavioral Disorders, 26*(1), 13-25.
- Frankel, K. A., Boyum, L. A., & Harmon, R. J. (2004). Diagnoses and presenting symptoms in an infant psychiatry clinic: comparison of two diagnostic systems. *J Am Acad Child Adolesc Psychiatry, 43*(5), 578-587.
- Glascoc, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities, 11*, 173-179.
- Gresham, F. M., & REACH Project. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children, 28*(4), 328-344.
- Guedeney, N., Guedeney, A., Rabouam, C., Mintz, A., Danon, G., Huet, M. M., et al. (2003). The Zero-to-Three diagnostic classification: A contribution of the validation of this classification from a sample of 85 under-threes. *Infant Mental Health Journal, 24*(4), 313-336.
- Harris-Murri, N., King, K., & Rostenberg, D. (2006). Reducing disproportionate minority representation in special education programs for students with emotional

- disturbances: Toward a culturally responsive response to intervention model. *Education and Treatment of Children*, 29(4), 779-799.
- Hill, L. G., Coie, J. D., Lochman, J. E., & Greenberg, M. T. (2004). Effectiveness of early screening for externalizing problems: issues of screening accuracy and utility. *J Consult Clin Psychol*, 72(5), 809-820.
- Kauffman, J. M. (2005). *Characteristic of Emotional and Behavioral Disorders of Children and Youth*. Columbus, OH: Jeffery W. Johnston.
- Kaufmann, R., & Perry, D. F. (2001, November 12, 2001). *Promoting social-emotional development in young children: Promising approaches at the national, state and community levels*. Paper presented at the The Kauffman Early Education Exchange Kansas City, MO.
- Landrum, T. J. (2000). Assessment for eligibility: Issues in identifying students with emotional or behavioral disorders. *Assessment for Effective Intervention*, 26(1), 41-49.
- Linan-Thompson, S., Vaughn, S., Prater, K., & Cirino, P. T. (2006). The response to intervention of English language learners at risk for reading problems. *J Learn Disabil*, 39(5), 390-398.
- Macy, M. G., Bricker, D. D., & Squires, J. K. (2005). Validity and reliability of a curriculum-based assessment approach to determine eligibility for Part C services *Journal of Early Intervention*, 28(1), 1-16.
- Macy, M. G., & Hoyt-Gonzales, K. (2007). A linked system approach to early childhood special education eligibility assessment. *Council for Exceptional Children*, 39(3), 40-44.

- Meisels, S. J. (1988). Developmental screening in early childhood: the interaction of research and social policy. *Annual Review of Public Health, 9*, 527-550.
- Minde, K., & Tidmarsh, L. (1997). The changing practices of an infant psychiatry program: The McGill experience. *Infant Mental Health Journal, 18*(2), 135-144.
- O'Connor, R. E., Harty, K. R., & Fulmer, D. (2005). Tiers of intervention in kindergarten through third grade. *J Learn Disabil, 38*(6), 532-538.
- Pretti-Frontczak, K. L. (2002). Using curriculum-based measures to promote a linked system approach. *Assessment for Effective Intervention, 27*(4), 15-21.
- Rich, B. A., & Eyberg, S. M. (2001). Accuracy of assessment: The discriminative and predictive power of the Eyberg Child Behavior Inventory. *Ambulatory Child Health, 7*, 249-257.
- Squires, J., Bricker, D. D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly, 16*, 405-419.
- Thomas, J. M., & Clark, R. (1998). Disruptive behavior in the very young child: diagnostic classification: 0-3 guides identification of risk factors and relational interventions. *Infant Mental Health Journal, 19*(2), 229-244.
- Thomas, J. M., & Guskin, K. A. (2001). Disruptive behavior in young children: What does it mean? . *Journal of American Academic in Children and Adolescent Psychiatry, 40*(1), 44-51.
- Van der Oord, S., Van der Meulen, E. M., Prins, P. J. M., Oosterlaan, J., Buitelaar, J. K., & Emmelkamp, P. M. G. (2005). A psychometric evaluation of the Social Skills

Rating System in children with attention deficit hyperactivity disorder. *Behaviour Research and Therapy*, 43, 733-746.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: a comparison of child and parent training interventions. *J Consult Clin Psychol*, 65(1), 93-109.

Weis, R., Lovejoy, M. C., & Lundahl, B. W. (2005). Factor structure and discriminative validity of the Eyberg Child Behavior Inventory with young children. *Journal of Psychopathology and Behavioral Assessment*, 27(4), 269-278.

Appendix Tables and Figures

Table 1 Background Characteristics of the Study Participants

Authors (year)	Purpose of the Study	Type of sample	No. of Subjects	Ages	Boys (%)	Race	SES
Screening Assessment							
Weis, Lovejoy, & Lundahl (2005)	Study 1: To examine the factor structure of the ECBI with a community sample of young children.	Community	489	2-6 yrs	53	C=86% A=10% O=4%	Low SES
	Study 2: To examine the discriminative validity of the ECBI with clinical-referred samples.	Clinical	115	4-6 yrs	67	C=68% (S2) A=32% (S2)	Low SES
Feil, Walker, Severson, & Ball (2000)	To assess the cross-cultural psychometric characteristics and validity of a multiple-gating screening procedure used by Early Screening Project (ESP) to screen and identify children at risk for behavioral problems in Head Start centers.	Community	Stage 1 954 (18% eligible for ECSE) Stage 2 126	3-4 yrs.	53	C=45% A=12% O=43%	Low SES
Feil & Becker (1993)	To assess the concurrent validity of the Walker/Severson Systematic Screening for Behavior Disorder (BBSD) in preschoolers	Preschools, 43% Head Start. Some with EBD, DD, LSD	Stage 1 121 Stage 2: 105	3-6 yrs	59	-	-
Squires, Bricker, Heo, & Twombly (2001)	To report the results of a psychometric study on the ASQ:SE.	Community	3,014 88 with EBD	3-63 months	50	C=59% A=9% O=32%	L1=20.6% L2=19.9% L3=22.8% L4=29.9% L5=6.8%
Rich & Eyberg (2001)	To assess the discriminative and predictive power of the ECBI.	Clinical	196, (98 w/ ODD, CD, or ADHD)	3-6 yrs	80	C=78% A=14% O=8%	L1=18% L2=27% L3=30% L4=11% L5=14%
Briggs-Gowen, Carter, Irwin, Wachtel, & Cicchetti (2004)	To examine the reliability and validity of the BITSEA	Community	1,237 (1 st year, 79 in home visits, 1,135 In 2 nd year)	12 to 36 months (<i>M</i> = 23.8, <i>SD</i> = 6.8)	49	C=66% A=16% O=18%	33.9 % blow or at board line (Median=\$50.600.)

Authors (year)	Purpose of the Study	Type of sample	No. of Subjects	Ages	Boys (%)	Race	SES
CBM Assessment							
Macy, Bricker, Squires (2005)	To determine if the AEPS:E accurately identified children eligible for early intervention services	Community	68 (33 EI, 35 non-EI)	18-36 months	54	C=74% A=6% O=20%	-
Bricker, yovanoff, Capt, & Allen, (2003)	To examine the usefulness of the AEPS to corroborate the findings of standardized, norm-referenced tests that establish eligibility for IDEA services for young children.	Community	861 (258 EI, 603 non-EI)	1-72 months	44	C=71% A=13% O=17%	-
Diagnostic Assessment							
Thomas & Clark (1998)	To examine the convergent validity of DC: 0-3	Clinics	64 (w/ disruptive behaviors concerns)	12-47 months	72	C=64% A=25% O=11%	-
Thomas & Guskin (2001)	To examine how the DC: 0-3 primary diagnoses, PIR-GAS, and CBCL/2-3 align with the DSM-III-R/IV and guide treatment.	Clinics	82 (w/ disruptive behaviors concerns)	18-47 (M = 36) months	77	C=71% A=23% O=6%	-
Frankel, Boyum, & Harmon (2004)	To demonstrate inter-rater reliability for diagnoses using both two diagnostic systems DSM-IV and DC: 0-3	Clinics	177 (166 w/ DSM-IV diagnosis; 132 w/ DC: 0-3 diagnosis)	0-58 months (M = 30.96)	56	C=73% A=10% O=17%	70% unknown, 28% <\$30,000 , 2% >\$30,000 .
Epstein, Nordness, Cullinan, & Hertzog (2002)	Study 1 To assess the long-term test-retest reliability of the Scale for Assessing Emotional Disturbance (SAED)	Schools	270	4-8 yrs (M = 5.7)	50	-	-
	Study 2 To examine the convergent validity of the SAED with children in kindergarten and first grade	Schools	123	5-8 yrs (M = 5.6)	52	-	-

Authors (year)	Purpose of the Study	Type of sample	No. of Subjects	Ag es	Boys (%)	Race	SES
Problem-solving model							
Fairbanks, Sugai, Guardino, & Lathrop (2007)	Study Investigating a RTI approach to behavior support in second-grade classrooms. This study focuses on evaluating a CICO (check in, check out) targeted intervention design.	School (2 classrooms)	10 (2 were in ECSE)	7-8 yrs	50	C=80% A=10% O=10%	-
	Study 2 To examine the effects of individualized function-based support for 4 participants whose behaviors did not respond as expected to the CICO intervention	School (1 classroom)	4	7-8 yrs	-	-	-
VanDerHeyden, Witt, & Barnett (2005)	To examine criteria for judging intervention response within the RTI model.	School (Math & Reading difficulties)	364	7-8 yrs	42	C=86% O=14%	-
O'Connor, Harty, & Fulmer (2005)	To examine the effects of second- and third-tier interventions delivered as needed from kindergarten through third grade on student's reading development and placement in special education by the end of third grade.	2 Schools (Reading problems)	31 (10 entered tier-2, 8 in tier-3)	-	-	C=68%	-
Linan-Thompson, Vaughn, Prater, & Cirino (2006)	To examine one aspect of RTI for English language learners (ELLs) who enrolled in either Spanish (SI) or English intervention (EI) programs: the incidence of students who did not respond to instruction even after a research-based, intensive, long-term intervention had been provided.	11 schools (7 SI, 4 EI, Reading problems)	SI: 64 (46 in tier-2) EI: 39 (29 in tier-2)	7-8 yrs	SI: 55 EI: 50	H=100	-

Note. Ethnicity: C = Caucasian, A = African American, O = Others, H = Hispanic. Instrument. ECBI = Eyberg Child Behavior Inventory; ASQ: SE = Ages and Stages Questionnaires; BITSEA = Brief Infant-Toddler Social and Emotional Assessment; AEPS: E = Assessment, and Programming System: Eligibility; DC: 0-3 Diagnostic Classification System: 0-3.

Table 2 Characteristics of Study Types and Methodology

Authors (year)	Design	Type of study	Data Collection Technique	Measurements	
				Instruments / Target behaviors	Raters
Screening Assessment					
Weis et al. (2005)	CoR	CA	Ratings, diagnostic interview, semi-structured behavior observation	<u>ECBI</u> : EP behaviors (Inattentive, ODD, CD, problem Behaviors) <u>DB-DRS</u> : Items correspond to DSM-IV-TR criteria for ADHD, ODD, and CD	Parents, psychologist, social workers
Feil et al. (2000)	Quesi-E	EI	<u>Tier 1</u> : Nomination and ranking <u>Tier 2</u> : Teacher and parent ratings <u>Tier 3</u> : Behavioral observations, parent rating	<u>CBCL (T2)</u> : IP & EP behaviors <u>SSRS (T2)</u> Social Skills (Cooperation, Assertion, Responsibility & Self-Control) <u>Peer Social Behavior observation code (T3)</u>	Teachers (T1, T2), parents (T2), & trained personals (T3)
Feil & Becker (1993)	Quesi-E	EI	<u>Tier 1</u> : Nomination and ranking <u>Tier 2</u> : Teacher ratings <u>Tier 3</u> : Behavioral observations	T2: <u>CEI-A & B</u> , Adaptive and Maladaptive Behavior Indexes, Behar, & Connors. T3: <u>SAET</u>	Teachers (T1, T2), trained personals (T3)
Squires et al. (2001)	CoR	TA	Questionnaires & ratings	<u>ASQ:SE</u> : Social Skills Related Behaviors <u>CBCL-2/3, 4-18</u> : IP & EP behaviors <u>Vineland SEECs</u> : Social-emotional competence <u>Parent Utility</u> : Satisfaction with the ASQ:SE.	Parents
Rich & Eyberg (2001)	CoR	CA	Structured interview, observation, videotape	<u>ECBI</u> : Conduct problems, Aggression, Inattention, <u>DSM-III-R</u> <u>ITSEA/BITSEA</u> : IP & EP, Dysregulation, Competence, Social Relatedness, Atypical behaviors, Maladaptive <u>CBCL/1.5-5</u> : IP & EP behaviors <u>MCDI-SF</u> : Vocabulary checklist <u>Parent Worry</u> : Social and emotional development, behavioral, and language <u>Evaluator ratings</u> : Problem & competence	Parents, clinicians
Briggs-Gowan et al. (2004)	CoR	TA	Questionnaire, videotape, parent interview	<u>ITSEA/BITSEA</u> : IP & EP, Dysregulation, Competence, Social Relatedness, Atypical behaviors, Maladaptive <u>CBCL/1.5-5</u> : IP & EP behaviors <u>MCDI-SF</u> : Vocabulary checklist <u>Parent Worry</u> : Social and emotional development, behavioral, and language <u>Evaluator ratings</u> : Problem & competence	Parents, Evaluators (home visit)

Authors (year)	Design	Type of study	Data Collection Technique	Measurements	
				Instruments / Target behaviors	Raters
CBM Assessment					
Macy et al. (2005)	CoR	CA	rating scales, Observation	<u>ASQ:SE:</u> Communication, motor (gross & fine), problem solving & personal social. <u>AEPS: E:</u> Motor (gross & fine), cognitive, adaptive, social-communication, & Social. <u>BDI/Gesell:</u> Personal social, adaptive, motor (gross & fine), Communication/language, and cognition (BDI only) <u>Vineland:</u> Motor, communication, daily living skills, socialization	Parent report for ASQ:SE & Vineland Trained personals finish AEPS:E, Gesell, BDI
Bricker et al. (2003)	CoR	CA	Rating scales, Observation	<u>AEPS: 0-3 & 3-6:</u> Motor (fine & gross), adaptive, cognitive, social-communication, social.	Trained personals
Diagnostic Assessment					
Thomas & Clark (1998)	CoR	TA	videotape, family interview, structure observation, videotape vignettes, ratings	<u>DC: 0-3:</u> 5 axes and PIR-GAS. <u>CBCL/2-3:</u> EP and IP behaviors <u>PCERA</u>	Clinicians (DC: 0-3) parents (CBCL/2-3)
Thomas & Guskin (2001)	CoR	TA	Videotape, family interview, diagnostic interview, structure observation, videotape vignettes, diagnostic measures	<u>DSM-III-R/IV:</u> emotional disorders, disruptive behaviors, adjustment disorders, other. <u>DC: 0-3</u> Axis I (primary diagnosis) and Axis II (relationship disorder), PIR-GAS. <u>CBCL/2-3:</u> EP & IP behaviors	Clinicians (DC: 0-3, DSM-III-R/IV) parents (CBCL/2-3)
Frankel et al. (2004)	CoR	TA	Retrospective Chart review, coding	<u>DC: 0-3:</u> 5 axes <u>DSM-IV</u>	Postdoctoral psychologists
Epstein et al. (2002)	CoR	TA	Teacher ranking, rating scales, observation	<u>SAED:</u> Inability to learn, Relationship Problems, Inappropriate Behavior, Unhappiness/Depression, Physical Symptoms/Fears (federal definition characteristics of ED). <u>SSBD</u> Critical Events, Maladaptive Behavior, Adaptive Behavior	Teachers

Authors (year)	Design	Type of Study	Data Collection Technique	Target behaviors	Tiers / Phase
Problem-Solving Models					
Fairbanks et al. (2007)	Qui-E (L)	PI	<u>Study 1:</u> Nomination, observation, ratings, referral records <u>Study 2:</u> Functional Checklist (FACTS), observation	EPs & academic engagement	<u>Study 1</u> (5 Phases): Baseline, CICO 70%, CICO 75%, CICO 80%, and CICO 90%. <u>Study 2</u> (6 Phases): Baseline, CICO 70%, CICO 75%, CICO 80%, FBA-based plan, and FBA-based plan-adjusted.
VanDerHeyden et al. (2005)	Qui-E (L)	PI	Coding, Observation, CBA, reading assessment, Teacher referral	Reading and Math difficulties	2 Tiers <u>T1:</u> Screening by four measures (CBM, CIBS-R, ITBS, teacher referrals) <u>T2:</u> CBA, school-wide screening, Protocol-based intervention.
O'Connor et al. (2005)	Qui-E (L)	PI	Reading measures (WRMT-R/NU)	Learning Disabilities	3 Tiers: <u>T1:</u> Professional Development (PD) for staffs; <u>T2:</u> Small-group Instruction (3 days per week); <u>T3:</u> Daily Small-Group or Individual Instruction
Linan-Thompson et al. (2006)	Qui-E (L)	PI	Reading Measures (WLPB-R)	Reading disabilities	2 Tiers <u>T1:</u> WLPB-R as screening tool, intervention provided for at-risk students. <u>T2:</u> WLPB-R was used to identify student's responsiveness toward intervention

NOTE.

Design: CoR = Correlational, Quasi-E = Quasi-Experimental, Quasi-E (L) = Quasi-Experimental Longitudinal

Type of Study: CA = Classification Accuracy, TA = Technical Adequacy, EI = Early Identification, PI = Prereferral Intervention

Table 3 Practice Characteristics Identified in the Studies

Authors (year)	Consensus Decision Making		functional Assessment			Adm. reliability		Pre-referral intervention	Linking	Utility
	Professional Consensus	Professional - parent consensus	Observation	Authentic Ecological Context	Multi-time/setting	Training	Inter-rater reliability			
Screening Assessment										
Weis et al. (2005)			x	x	x					
Feil et al. (2000)			x	x	x	x	x			
Feil & Becker (1993)			x	x	x	x	x			
Squires et al. (2001)			x	x	x					
Rich & Eyberg (2001)			x			x	x			
Briggs-Gowan et al. (2004)			x	x	x		x			
CBM Assessment										
Macy et al. (2005)			x	x	x	x	x			x
Bricker et al. (2003)			x	x	x	x				x
Diagnostic Assessment										
Thomas & Clark (1998)	x	x	x	x	x	x		x	x	x
Thomas & Guskin (2001)			x	x		x				x
Frankel et al. (2004)						x	x			
Epstein et al. (2002)			x	x	x	x			x	x
Problem-Solving Models										
Fairbanks et al. (2007)	x		x	x	x	x	x	x	x	x
VanDerHeyden et al. (2005)	x		x	x	x	x	x	x	x	x
O'Connor et al. (2005)	x			x	x	x		x	x	x
Linan-Thompson et al. (2006)					x			x	x	x
Percentage of Studies (%)	25%	6.3%	81.3%	81.3%	81.3%	75%	50%	31.3%	37.5%	56.3%

Table 4 Major Findings and Results Reported in the Studies

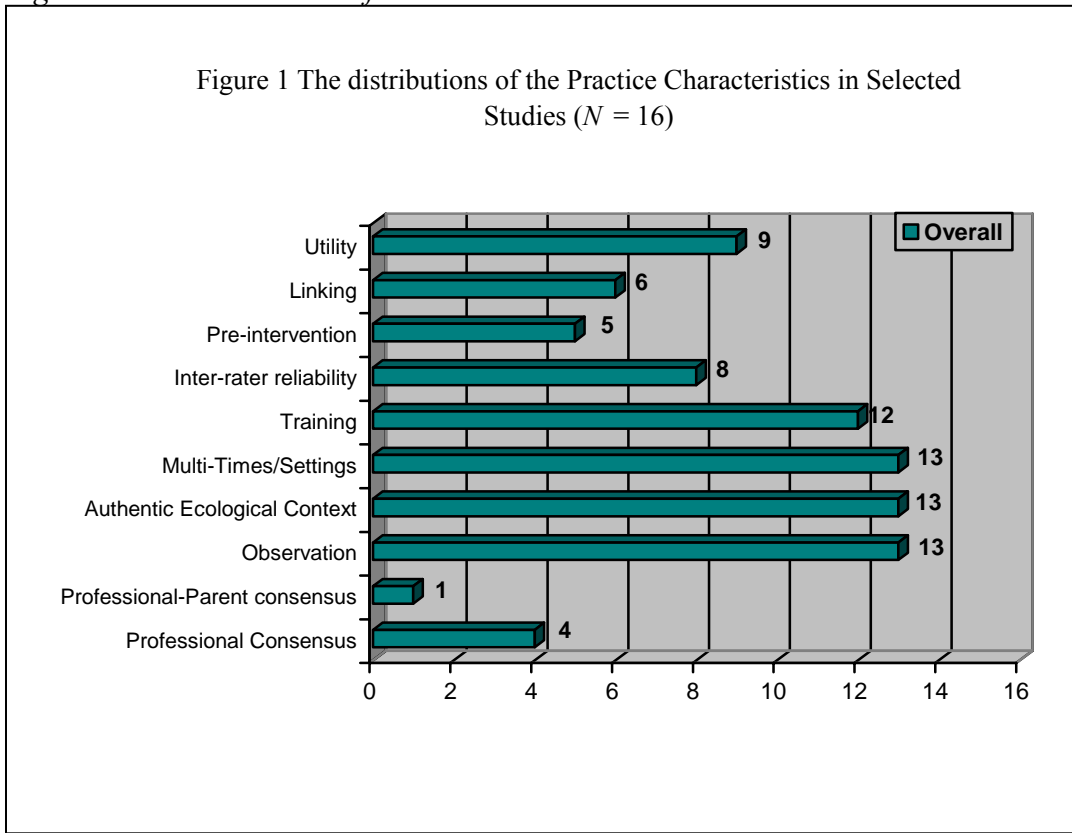
Authors (year)	Major Findings / focus on statistics writing	Results
<u>Screening Assessment</u>		
Weis et al. (2005)	<p>Study 1</p> <p>1. CFA analysis found that three-factor model provided a significantly better fit than one- and two-factor models ($\Delta X^2(2, N=489) = 249.76, p < .01$). Three factors are Inattentive, Oppositional Defiant Behavior, and Conduct Problem Behavior.</p>	<p>1. The three-factor model provided a significantly better fit than either a one-factor or two-factor model. These results indicate that even at a young age, symptoms of inattention, oppositional behavior, and conduct problems can be differentiated on the ECBI.</p>
	<p>Study 2 - <u>Predictive Accuracy</u></p> <p>1. ECBI show fair sensitivity for Inattentive (77%) and ODD (75%), but poor for Conduct Problem (63%).</p> <p>2. Specificity was excellent ranged from 91% to 94%. Positive Predictive Value (PPV) was good for Inattentive (85%) and ODD (80%), but poor for CP (63%).</p>	<p>1. Scores above the 90th percentile cutoff might be used to identify children with significant inattentive and ODD symptoms.</p> <p>2. The ECBI components may be useful to screen community-based and clinical-referred young children for externalizing symptoms, but may be less useful to identify children with specific behavior problems.</p>
Feil et al. (2000)	<p><u>Psychometric Findings</u></p> <p>1. Correlation coefficients among the ESP, CBCL and SSRS showed good agreement between measure, ($r = .44$ to $.91$); moderate agreement across raters ($r = .38$) in SSRS Externalizing), and an overall better relationship among externalizing subscales.</p>	<p>1. In Stage Two, the addition of the direct observations resulted in fewer children exceeding cut-off criteria, which would seem to increase sensitivity. While the ESP Stage Two teacher measures identify a larger percentage of children, the Stage Three Observations show increased specificity.</p>
	<p><u>Others:</u></p> <p>1. No significant differences in the number of referrals when using the ESP among ethnic groups.</p> <p><u>Productivity Accuracy</u></p> <p>1. Sensitivity was 97% for externalizers group and 100% for internalizers group; Specificity was 90% for externalizers group and 97% for internalizers group).</p>	<p>1. The low false positive and low false negative rate illustrates the potential utility of the screening system, giving evidence that behavior problems may be identified accurately among preschool children.</p>
Feil & Becker (1993)	<p><u>Psychometric Findings</u></p> <p>1. PSBP shows good results in the psychometric characteristics, including interrater reliability ($r = .55$ to $.83$), test-retest reliability ($r = .49$ to $.91$), concurrent validity in preschool populations ($r = .25$ to $.84$).</p>	<p>2. The consistency across measures illustrates the potential utility of the screening system, giving evidence that behavior problems may be identified accurately among preschool children.</p>

Authors (year)	Major Findings / focus on statistics writing	Results
Squires et al. (2001)	<p><u>Psychometric Findings</u></p> <p>1. Internal consistency ranged from .67 to .91, with an overall alpha of .82. Test-retest reliability was 94% ($N=344/367, p<.05$).</p> <p><u>Predictive Accuracy</u></p> <p>1. Sensitivity was 81.6% and Specificity was 91.5%. Percentage agreement across measure was 89%. Under referral rate was 2.4% and overreferral rate was 7.7%.</p> <p><u>Others</u></p> <p>2. Over 96% of parents indicated that ASQ: SE items were easy to read and understand.</p>	<p>1. The ability of the ASQ: SE to detect atypical social-emotional development (sensitivity) or the ability of the ASQ: SE to correctly identify typically developing children remained high.</p>
Rich & Eyberg (2001)	<p><u>Productivity Accuracy</u></p> <p>1. The Sensitivity was 96%; the specificity was 87%; and the hit rate was 91%. When the sample was weighted so as to represent a 20% prevalence estimate of DB in young children, the sensitivity was 90%, the specificity was 87%, and the hit rate was 88%.</p>	<p>1. All results support the use of the ECBI for screening young children with disruptive behavior problems for referral for comprehensive psychological evaluation and treatment.</p>
Briggs-Gowan et al. (2004)	<p><u>Technical adequacy</u></p> <p>1. Test-retest reliability was excellent ($\alpha = .85$ to $.87$) and interrater agreement was good (correlation coefficient = $.61$ to $.68$). Supporting validity, BITSEA problems correlated with concurrent evaluator problem ratings and CBCL/1.5-5 scores and also predicted CBCL/1.5-5 and ITSEA problem scores one year later.</p> <p><u>Predictive accuracy</u></p> <p>1. Combined BITSEA problem/competence cut points identified 85% of sub-clinical/clinical CBCL/1.5-5 scores, while maintaining acceptable specificity (75%).</p>	<p>1. Findings support the validity of the BITSEA, given the established success of the CBCL in discriminating referred and nonreferred children. In addition, findings also support the BITSEA as a screener for social-emotional / behavioral problems and delays in social-emotional competence.</p>
CBM Assessment		
Macy et al. (2005)	<p><u>Predictive Accuracy</u></p> <p>1. The psychometric values between AEPS: E and BDI/Gesell show that the sensitivity of AEPS: E was 100%, the specificity was 89%. The percent agreement was 94%.</p> <p><u>Psychometric findings</u></p> <p>1. AEPS: E overall scores was significantly correlated with BDI overall scores ($r = .65$), Vineland overall scores ($r = .52$), and subscales in Gesell ($r = .42$ to $.62$). Interrater agreement between two trained observers who scored the AEPS: E was from $.80$ to $.89$.</p>	<p>1. This study was an initial investigation of the accuracy of the AEPS: E in correctly classifying toddlers who were eligible for Part C services. Findings suggest that the AEPS: E correctly classified all eligible children and most ineligible children. Results lend tentative support to the potential usefulness of the AEPS: E to determine the eligibility of toddlers for specialized services.</p>

Authors (year)	Major Findings / focus on statistics writing	Results
Bricker et al. (2003)	<p><u>Predictive Accuracy</u></p> <p>1. The overall sensitivity for AEPS was 88.67% for birth to three, 90.17% for three to six; the overall specificity was 77.8% for birth to three, 75% for three to six; the Under identification was 2.8% for birth to three, 3% for three to six; The over identification was 15.7% for birth to three, 19.3% for three to six.</p>	<p>1. The sensitivity of the AEPS cutoff scores for both levels is consistently high, suggesting that the cutoff scores classify children appropriately as eligible. Overall, the findings provide substantial support for the use of the AEPS as a tool to corroborate eligibility decisions.</p>
<u>Diagnostic Assessment</u>		
Thomas & Clark (1998)	<p>1. 84% (n=64) have disruptive behavior as a main concern. Of 64 children, the primary DC: 0-3 diagnoses were TSD (23%), DA (14%), RD (30%); while the four most common DSM-IV diagnoses were Adjustment Disorder (30%), ODD (14%), Dysthymia (14%), and ADHD (13%). The three major DC: 0-3 diagnostic groups differed minimally when examining CBCL child symptoms alone (n = 49).</p> <p>2. Lower scores on the PIR-GAS were marginally associated with higher level of child aggression on the CBCL (p<.06).</p>	<p>1. DC: 0-3 Axes I and II and the PIR-GAS provide diagnostic tools that help to differentiate this heterogeneous group of young children through identification of specific risk factors that guide intervention.</p> <p>2. DSM-IV's purely descriptive diagnoses do not adequately describe the difficulties of these very young children who, in response to a broad array of risk factors, display a small repertoire of salient disruptive behaviors.</p>
Thomas & Guskin (2001)	<p>1. Thirty-four percent of children (n = 28) received a DSM diagnosis of disruptive behavior disorder/ADHD (n= 14), ODD (n = 11), Disruptive behavior (n = 3); while in DC: 0-3 system, children received a diagnosis of RD (n = 16), DA (n = 10), and TSD (n = 2). Of DA children, 41.2% had relationships categorized as disordered, contrasting with 21.1% of those with TSD, and 10.3% of those with RD.</p>	<p>1. Using the DSM disruptive disorder/ADHD descriptions alone, we might easily have missed the specific constitutional/maturational, affective, or traumatic risk factors identified by DC: 0-3.</p> <p>2. Parent-child relationship is an important risk factor associated with the presentation of internalizing and externalizing symptoms.</p>
Frankel et al. (2004)	<p>1. Reasonable interrater reliability was observed for DSM-IV and DC: 0-3 (averaged 89.9% agreement).</p> <p>2. The major discrepancy between the two diagnostic systems involved the DC: 0-3 diagnostic category of "regulatory disorders."</p>	<p>1. There is a good concordance between DSM-IV and DC: 0-3 diagnoses where the systems overlap. DSM-IV disruptive behavior disorders may be better conceptualized in DC: 0-3 as regulatory disorders, leading to alternative conceptualization of the disorder and a different course of treatment.</p>
Epstein et al. (2002)	<p><u>Study 1</u></p> <p>1. For each of the SAED subscales, 7 months interval reliability showed a moderate to high correlation (r = .051 to .84, p<.001).</p>	<p>1. Teachers and other decision makers can accurately identify areas of emotional or behavioral concern for general education students with the SAED and that those areas of concerns that were identified by the SAED are stable over time.</p>

Authors (year)	Major Findings / focus on statistics writing	Results
Epstein et al. (2002)	<p><u>Study 2</u></p> <p>1. Correlations between SAED and SSBD were found to be low, middle, or high across the subscales. Most (16 of 18) of the correlations were statistically significant $r = -.238$ to $.810$ ($p < .01$ to $.001$).</p>	<p>1. Correlations were strong between subscales that assessed similar constructs between SAED and SSBD.</p>
<i>Problem-Solving Models</i>		
Fairbanks et al. (2007)	<p>1. All 10 students experienced the CICO intervention. Of the 6 students whose behaviors were not responsive to CICO, 4 students received more individualized function-based interventions.</p> <p>2. The ODR rate per day was decreased from .85 to .41 after CICO intervention was implemented.</p> <p>3. After implementing CICO, teachers rated problem behavior as a 3 or lower comparing to 4 or higher before CICO.</p>	<p>1. A slightly more intensive but efficient intervention (i.e., CICO) was effective in supporting the behavioral success of four students whose problem behaviors were initially unresponsive to general classroom management practices.</p> <p>2. Teachers reported that the interventions were easy to implement and improved the general climate of the classroom, and that being on CICO or an individualized plan was viewed as a positive experience by students.</p>
VanDerHeyden et al. (2005)	<p><u>Predictive accuracy</u></p> <p>1. Overall agreement between ITBS and CBA was 96%, the sensitivity was 33%, specificity was 99%, the positive predict value was 50%, and the negative predictive power was 97%.</p> <p><u>Others</u></p> <p>2. The slope of intervention data was positively correlated with the slope of the generalization data for math and reading (50% to 53%).</p>	<p>1. Brief instructional group and individual sessions may assist with decision making.</p> <p>2. RTI presents definitional "openness," procedural, sampling, and measurement variations, process challenges, and many intervention variables.</p>
O'Connor et al. (2005)	<p>1. Following 4 years, the rate of placement for special education was 8% comparing to 15% in the historical control group.</p> <p>2. Nearly 40% of students who received Tier 3 progressed sufficiently to maintain average performance through out third grade without additional interventions.</p>	<p>1. There was much more "in-and-out" movement than expected. A model of tiered intervention appears to help students who struggle with reading acquisition and could help to document alternative instruction.</p>
Linan-Thompson et al. (2006)	<p>1. In SI group, at the end of 1st grade, 97% of students met RTI criteria in the Intervention group comparing to 70% in control group. At the end of 2nd grade, 100% met RTI criteria in the intervention group comparing to 92% in the control group.</p> <p>2. In EI group at the end of 1st grade, 91% of students met RTI criteria in the intervention group comparing to 42% in control group. At the end of 2nd grade, 94% met RTI criteria in the intervention group comparing to 44% in the control group.</p>	<p>1. Overall, the majority of the intervention students responded to the intervention and maintained that responder status to the end of second grade.</p>

Figure 1 The distributions of the Practice Characteristics in Selected Studies



List for selected articles

1. (Weis et al., 2005)
2. (Feil et al., 2000)
3. (Feil & Becker, 1993)
4. (Squires et al., 2001)
5. (Rich & Eyberg, 2001)
6. (Briggs-Gowan et al., 2004)
7. (Macy et al., 2005)
8. (Bricker et al., 2003)
9. (Thomas & Clark, 1998)
10. (Thomas & Guskin, 2001)
11. (Frankel et al., 2004)
12. (Epstein et al., 2002)
13. (Fairbanks et al., 2007)
14. (Van der Oord et al., 2005)
15. (O'Connor, Harty, & Fulmer, 2005)
16. (Linan-Thompson, Vaughn, Prater, & Cirino, 2006)