



Cornerstones

Practice-Based Research Syntheses of Child Find, Referral, Early Identification, and Eligibility Practices and Models
Volume Two, Number Three

November 2006

Research Foundations for Using Clinical Judgment (Informed Opinion) for Early Intervention Eligibility Determination

Stephen J. Bagnato
Janell Smith-Jones
Margaret Matesa
Eileen McKeating-Esterle

Clinical judgment, or informed opinion, refers to the knowledgeable perceptions of caregivers and professionals about the elusive and subtle capabilities of children in different settings that must be defined and quantified so that individuals or teams are able to reach accurate decisions about eligibility for early intervention. The characteristics of clinical judgment examined in 20 research studies highlight the practices that best enable individuals to describe child-related information and that contribute to effective evaluation decisions. Our findings identified five practice characteristics relevant to the clinical judgments by individuals and one for the clinical judgments by teams: (1) an operational definition of the child characteristics to be judged, (2) a structured format for quantifying those characteristics, (3) information derived from multiple settings and individuals, (4) training in methods that structure and quantify characteristics, and (5) decision making based on a consensus of information from multiple settings and individuals. Implications for eligibility determination for Part C early intervention and for further research are described.

Purpose

Clinical judgment, or informed opinion, refers to the knowledgeable perceptions of caregivers and professionals about the elusive and subtle capabilities of children in different settings that must be defined and quantified so that individuals or teams are able to reach accurate decisions about eligibility for early intervention. Clinical judgment is considered most effective when it involves measurement and decision-making strategies that integrate and synthesize information about multiple aspects of children's behavior and development.

The purpose of this practice-based literature review and research analysis is to examine available evidence concerning the use of clinical judgment as a process for documenting, integrating, and evaluating child-related information used to determine eligibility for Part C early

intervention. It was conducted to discern whether available evidence warrants the use of clinical judgment, or informed opinion, as an eligibility determination practice as indicated in IDEA (Individuals with Disabilities Education Act, 20 U.S.C. § 1432(5), 1435(a)(1), 1997). More specifically, we examined the research-based practice characteristics of clinical judgment associated with accurate, representative, and reliable assessment out-

This *Cornerstones* is a publication of the Tracking, Referral, and Assessment Center for Excellence (TRACE) funded by the U. S. Department of Education, Office of Special Education Programs (H324G020002). The opinions expressed are those of the authors and do not necessarily reflect the views of the U.S. Department of Education. TRACE is an initiative of the Center for Improving Community Linkages, Orelena Hawks Puckett Institute (www.puckett.org). Copyright © 2006 by the Orelena Hawks Puckett Institute. All rights reserved.

comes applicable to eligibility determination and diagnosis. The literature review and analysis was conducted as part of research at the Tracking, Referral and Assessment Center for Excellence (Dunst & Trivette, 2004; Dunst, Trivette, Appl, & Bagnato, 2004).

Background

Meehl's (1954) seminal work on clinical decision-making strategies touched off a flurry of research on the reliability and validity of clinical judgment (Westen & Weinberger, 2004). In the past 50 years, the effectiveness of clinical-judgment strategies in decision making has been used in the fields of education, medicine, mental health, business, and criminal justice. Early research findings (Dawes, Faust, & Meehl, 1989; Garb, 1994; Holt, 1970; Marchese, 1992; Meehl, 1954; Sawyer, 1966; Sines, 1970; Wiggins, 1981) suggested that statistical or actuarial prediction strategies were slightly more accurate than clinical methods. More recent research indicates that both strategies make valuable contributions and are perhaps complementary approaches to measurement and decision making (Ægisdóttir et al., in press; Grove, Zald, Lebow, Snitz, & Nelson, 2000).

Despite limited research evidence, clinical judgment is widely practiced and valued by early intervention practitioners (Fleischer, Belgredan, Bagnato, & Ogonosky, 1990; Hayes, 1990; Hunt, Johnson, Owen, Ormerod, & Babbitt, 1990; LeLaurin, 1990; LeVan, 1990; McCloskey, 1990; Miltenberger, 1990; Sexton, Thompson, Perez, & Rheams, 1990; Short & Simeonsson, 1990). This appears to be the case because it safeguards against eligibility being based on isolated information or test results alone and is particularly important for eligibility decisions when the use of standardized measures are not useful options (Shackelford, 2002). Some early intervention studies, which examined clinical judgment for early detection and classification of young children with delays or disabilities, revealed promising avenues for future research on clinical judgment (Bagnato & Neisworth, 1999; Glascoe, 1991; Kochanek, Kabacoff, & Lipsitt, 1990; Records & Tomblin, 1994; Sampers, Cooley, Cornelius, & Shook, 1996). Much of this work focused on the use of assessment tools because these permit individuals to quantify their judgments about the elusive and subtle aspects of children's behavior and development in different environments. Several measures have been used to assess broad areas of functioning, including overall developmental status, health, and family factors (Henderson & Meisels, 1994; Kochanek et al., 1990), while others have been used to assess specific aspects of health and development, including physical robustness (Casey, McIntyre, & Leveno, 2001), cognitive and perceptual skills (Gresham, Reschly, & Carey, 1987), motor

skills (Sampers et al., 1996), speech and language skills (Glascoe, 1991; Records & Tomblin, 1994), and temperament and self-regulation (Bagnato & Neisworth, 1999).

One of the essential elements of clinical judgment is that they structure and quantify perceptions, observations, and judgments through ratings and classification formats. Some clinical-judgment processes and instruments (rating scales, observation formats, and functional classification systems) have yielded adequate reliability, validity, and utility for the purpose of decision making regarding detection, diagnosis, treatment planning, and progress monitoring, plus greater sensitivity than conventional tests about the subtle attributes of children (e.g., muscle tone, alertness, reaction to novelty), particularly those with complex disabilities (Bagnato, 1984; Bagnato & Neisworth, 1990; Bagnato, Neisworth & McClosky, 1994; Bailey, Buysee, Simeonsson, Smith, & Keyes, 1995; Bailey, Simeonsson, Buysse, & Smith, 1993; Glascoe, 1991; Gresham et al., 1987; Sampers et al., 1996; Simeonsson, Bailey, Smith, & Buysee, 1995; Simeonsson, Huntington, Short, & Ware, 1982; Suen, Lu, Neisworth, & Bagnato, 1993; Suen, Logan, Neisworth, & Bagnato, 1995).

Description of the Practice

The studies included in this literature review and analysis were appraised using five practice characteristics that were considered useful for: (a) identifying children at risk for or with disabilities, (b) documenting the needs of children enrolled in early intervention services, or (c) effectively using judgment-based scales and formats to assess children's skills and progress. The research-based practice characteristics include: (1) an operational definition of the specific child characteristics, attributes, skills, and support needs that are being judged; (2) a structured format for quantifying those characteristics; (3) the collection of information derived from multiple settings and individuals; (4) training in methods that structure and quantify children's characteristics, traits, skills, and support needs; and (5) decision making based on a consensus of information from multiple settings and individuals. Studies included in this paper were evaluated in terms of the use of these five practice characteristics. The interested reader is referred to Schalock and Luchasson (2005) for other characteristics of clinical judgment.

Operational definitions were examined in terms of researchers' use of clear, concise, and observable constructs (e.g., expressive language skills, muscle tone, articulation skills) measured by clinical-judgment scales. Operational definitions help ensure that different individuals have the same understanding of what is being rated, so that they can judge these dimensions through similar lenses. This was considered especially important when evaluating more elusive aspects of development such as

self-regulation and temperament.

Structured opinions (rating formats) were examined to further establish the specific aspects of development being assessed and the rating criteria being used as part of an assessment. They were examined by pre-established, standard formats that included written scales, interviews, archival coding schemes, and the use of quantitative or qualitative rating criteria.

Multiple informants and sources of information. These practice characteristics were examined by the extent to which studies pooled information about children's functioning from a variety of sources (assessments, medical records, etc.) and by a variety of people in the decision-making process. This practice characteristic underscores the importance of using comprehensive information for educational and clinical decision making. It involves gathering information about children's functioning in the home, at school, and in other settings such as doctors' or therapists' offices.

Consensus decision making. This practice characteristic was examined in order to determine whether information was both taken from more than one individual and/or setting and considered collectively in decisions about children's developmental status, progress, and/or needs.

Training to guide and facilitate reliable ratings. The use of training was examined to determine whether the raters had some level of training in completing clinical-judgment measures. Training ranged from overviews of rating intent and instructions to the achievement of rater consensus.

Search Strategy

Search Terms

The following search terms were used to identify relevant published and unpublished position papers, literature reviews, and research studies: *clinical judgment, informed clinical opinion, team decision-making, judgment-based assessment, subjective judgment, and qualitative assessment.* The search was done broadly in the fields of psychology, medicine, allied health, developmental disabilities, special education and early intervention.

Sources

The primary databases searched were: Psychological Abstracts online (PsycINFO), Social Sciences Citation Index (SSCI), Education Resource Information Center (ERIC), Ovid, MEDLINE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Health Source: Nursing/Academic Edition, Cochrane databases, and Medscape. Additionally, we conducted both a secondary search of an Endnote database and selective searches of unpublished master's theses and doctoral dissertations.

Selection Criteria

No studies were found that examined the use or effectiveness of clinical-judgment processes or instruments specifically for eligibility determination. Therefore, the utility of clinical judgment for early intervention was examined in terms of research on differential diagnosis and congruence among individuals in determining the extent of functional problems in development or behavior. Studies were included in the review if the investigation: (a) involved the clinical judgments of professionals or parents about the diagnosis, developmental status, and/or progress of children, particularly children under 5 years of age, (b) reflected one or more moderate to strong estimates of diagnostic outcomes, interrater agreement on developmental outcomes or significant contributions to estimates of developmental status, or (c) involved at least a subsample of children with developmental delays or disabilities.

Search Results

Twenty studies from the fields of early intervention, special education, early childhood education, speech/language pathology, and medicine were identified from the literature search that met the selection criteria. Table 1 shows selected characteristics of the study participants.

Participants

The studies included 157,681 participants that ranged in age from birth to 17 years of age. The largest majority of these participants were from a single study (Casey et al., 2001). Most studies included children with developmental delay, mental retardation, or other disabilities (Fragile X syndrome, cerebral palsy, spina bifida, autism, behavioral/emotional disorders, learning disability, visual, sensory, motor, and speech/language impairments). Several studies also included children without identified disabilities or delays (Bagnato & Neisworth, 1990; Glascoe, 1991; Gresham et al., 1987; Henderson & Meisels, 1994; Kochanek et al., 1990; Records & Tomblin, 1994; Sampers et al., 1996; Suen et al., 1993; Suen et al., 1995).

Type of Study

The studies of clinical-judgment practices were divided into three types: Early detection and classification (N = 6), parent-professional congruence (N = 6), and judgment-based scales and formats (N = 8). The first set of studies focused on the effectiveness of clinical assessment tools to accurately predict developmental delays or disabilities. The second set examined the extent to which judgment-based scales facilitated agreement between parent and professional ratings of children's abilities. The third group of studies examined the effectiveness of

judgment-based scales in providing accurate estimates of child outcomes.

Research Designs

The 20 studies included 11 case studies, 2 generalizability studies, and 7 cohort or longitudinal studies (see Table 3). The longitudinal studies incorporated follow-up assessments with two to six time points. The length of time between initial and follow-up assessments ranged from 28 days to 7 years.

Outcomes

Study outcomes were examined in terms of (1) *diagnostic accuracy* (i.e., sensitivity or specificity), (2) *agreement rates* between clinical-judgment tools and other early childhood measures and *interrater agreement levels* on clinical-judgment measures, and (3) *statistical contributions of multiple perspectives* to the assessment process (i.e., reliability and standard errors of measurement values associated with using multi-source information).

Results

Practice Characteristics

Table 2 shows the focus of and the extent to which each characteristic was included in each study. Judgment-based assessment strategies used within this set of studies incorporated at least two of the five practice characteristics examined in this analysis.

Operational definitions. All 20 studies included operational definitions of the areas being appraised through clinical-judgment strategies. Studies were selected for inclusion in the review because they employed clinical-judgment strategies that structured the collection of children's information in ways that were clear and organized. Within this body of research, consistent evidence for the use of this particular characteristic was examined. Beyond the theoretical construct being assessed (e.g., early language skills, early motor skills), operational definitions provide structure at the most basic level of measurement by specifying the particular dimensions or traits under consideration. This helps ensure that different individuals assign their ratings to the attributes under consideration using similar lenses.

Structured opinions (rating formats). Structuring the rating of informed opinion in conjunction with using operational definitions further ensures reliable assessment ratings by guiding how specific attributes are quantitatively or qualitatively rated. Most studies ($N = 14$) used preestablished formats to structure or quantify ratings of clinical or educational information. Rating formats varied in terms of mode (e.g., written scales, interviews, archival data coding), rating scheme (e.g., qualitative and

quantitative scales, open-ended responses), and level of detail. The majority of studies used written scales to assess children's abilities for diagnostic, progress monitoring, and/or educational planning purposes.

Multiple informants and sources of information. This practice characteristic underscores the importance of using different informants and sources of information, gathering information about children's functioning in the home, at school, and/or as part of evaluative situations such as doctor's or therapist's visits. Parents, teachers, psychologists, and speech, occupational, and physical therapists bring unique perspectives given the specific nature of their assessments. Considered together, their contributions create more complete pictures of children's strengths and needs, and in turn more complete information for eligibility decision making.

Fifteen of the 20 studies (75%) included procedures for gathering information on child competencies in different settings and by different individuals, and these procedures were found to increase the accuracy of diagnostic outcomes (see below). Information was most often gathered from a combination of clinical-judgment tools, curriculum-based measures, traditional assessments, and/or clinical impressions in six of the nine studies and involved at least three raters. Furthermore, assessments were most often completed by parents, teachers, and one other developmental specialist (e.g., psychologist, speech/ language therapist, diagnostician). The use of two assessment opinions was typically associated with interdisciplinary team decisions where teams were typically composed of a combination of parents, teachers, therapists, psychologists, and/or medical personnel.

Consensus decision making. Consensus decision making entails the integration and synthesis of information from different persons and sources to facilitate appropriate diagnostic and planning decisions. Ideally, information is obtained from four or more individuals (Suen et al., 1993) where informational sources include medical records, educational records, prior assessments, and other sources. Six studies examined the extent to which individuals used collective assessment information to make comprehensive or accurate assessment decisions. In every case, integrating all (as opposed to some) informational sources provided the most reliable and accurate snapshots of children's development. Consensus decision-making processes were part of the measurement strategies and procedures in 60% of the studies and were found to be effective in producing the most accurate and representative diagnostic results.

Training to guide and facilitate reliable ratings. Few studies included this practice characteristic, while position papers (Danaher, Shackelford, & Harbin, 2004; Hemmeter, Joseph, Smith, & Sandall, 2001; Sandall, McLean, & Smith, 2000) regard it as essential. In fact,

training of raters was described for only four tools (Perceptions of Developmental Skills, System to Plan Early Childhood Services, Movement Assessment Inventory, and APGAR). The scope of trainings included brief overviews of rating intent and instructions, achievement of rater consensus, and formalized training sessions. Training was described in 45% of the studies, but only two studies focused on the importance of this dimension in obtaining accurate results.

Effectiveness of Clinical Judgment

Table 3 shows the study types, results, and major findings. Seven studies examined the effectiveness of clinical-judgment measures to detect children at risk for or with a delay or disability. Researchers calculated sensitivity indexes or percentages of children correctly identified as delayed or disabled and specificity indexes or percentages of children correctly classified as nondisabled. Overall, measures were found to be accurate both in identifying children with delays or disabilities and in identifying children without disabilities. Sensitivity ranged from 72% to 94%; where most indices were 80% or higher. One study predicted neonatal death rates based on APGAR score ranges. Low APGARs (0–3) were associated with a significant number of neonatal deaths (24%–32%) and high scores (≥ 7) were associated with extremely low neonatal death rates (<1%).

Kochanek and his colleagues (1990) examined the importance of multiple sources of information in the prediction of disabilities. Child performance and demographic variables were evaluated with logistic regression analyses. Researchers reported moderate to strong sensitivity indexes, which ranged between 62% and 92%, but low to moderate specificity indexes which ranged between 30% and 74%. Outcomes varied depending on the variable groupings addressed within the analyses.

Congruence Across Measures and Between Raters

A few studies examined the extent to which clinical-judgment tools provided similar results to norm-referenced and curriculum-based measures. Bagnato and his colleagues (Bagnato, 1984; Bagnato & Neisworth, 1985) found moderate to strong levels of agreement (reliability coefficients ranged from $r = .65$ to $.89$) comparing ratings of clinical-judgment tools with those of norm-referenced and curriculum-based measures. Despite being based on a small number of studies, the evidence suggests that outcomes measured by clinical-judgment tools can yield results similar to those measured by norm-referenced and curriculum-based measures.

Several research teams (Blacher-Dixon & Simeonson, 1981; Gradel, Thompson, & Sheehan, 1981; Records & Tomblin, 1994) examined the extent to which

different raters' measurements were consistent and found considerable variability in coefficients of agreement ranging from very weak to very strong ($r = .24$ to 1.00). In some instances, agreement between mothers' and teachers' ratings varied. Overall, researchers found them to be consistent and addressed the importance of including multiple perspectives in children's assessments. Examining this idea in further detail, Suen and colleagues (1993, 1995) measured the degree to which information from multiple raters improved the accuracy of developmental appraisals. Findings indicated that as the number of raters increased from two to four, reliability estimates increased and standard errors of measurement decreased. Results also suggested that information from both caregivers and professionals provide the best representation of children's skills and needs when we recognize that each person has relevant and real, but incomplete information.

Conclusion

Two questions were addressed in this practice-based literature review and analysis: (1) Can informed clinical judgment be used effectively for assessment and diagnostic purposes? and (2) what practice characteristics are essential for accurate clinical judgments? Results indicated that it is possible to structure and quantify observations to make clinical judgments so that reliable, valid, and useful information for early detection is obtained. However, caution is warranted given the limited amount of evidence and the fact that relevant studies focused more on the status and progress of children already enrolled in early intervention programs as opposed to the status of children being assessed for potential enrollment in these programs.

Five practice characteristics were identified as important components of informed clinical judgment: an operational definition of the specific child characteristics, attributes, skills, and support needs that are being judged; a structured format for quantifying those characteristics; the collection of information derived from multiple settings and individuals; training in methods that structure and quantify children's characteristics, traits, skills, and support needs; and decision making based on a consensus of information from multiple settings and individuals. These characteristics provide a framework for individuals and teams, especially when rating challenging or elusive traits, to assess children's strengths and needs in a reliable, comprehensive, and meaningful fashion. The information reported in this paper is a first step, as more well-designed studies are needed to validate the professional confidence placed in clinical-judgment strategies and to ensure their valid use by early intervention practitioners.

Evidence to support the use of clinical-judgment tools for measuring children's needs and skills by using multiple sources of information in therapeutic or educational decisions provided the basis on which the study findings were evaluated. Overall, the use of these practice characteristics in studies found to exemplify sound clinical-judgment strategies and solid outcomes demonstrates some support for the use of operational definitions, structured formats, and multi-source information; moderate support for the use of consensus decision-making processes; and inconsistent support for assessment training. Available evidence suggests that clinical judgment or informed opinion holds promise as a potentially effective strategy for use in early intervention for eligibility determination with the proper use of the identified practice characteristics.

Implications for Practice

Evidence indicates that informed clinical judgment can provide reliable and valid assessment outcomes, but questions remain as to what is needed to make it a more integral part of eligibility determination. Perhaps important first steps are (1) knowing when it is an option and (2) knowing when it should be used. In states where informed clinical opinion is used for eligibility determination, awareness of the regulations and knowledge of when it is most appropriate are essential. As stated in the Individuals with Disabilities Education Improvement Act (2004), informed clinical opinion is particularly important when standardized measures are inappropriate or unavailable. It can be used to detect abilities in children who are inconsistent in or of a low threshold in their expression. Furthermore, clinical judgment can unify and facilitate team decisions about child characteristics and specific programmatic and intervention needs and enable parents to be active and integral participants in the assessment process.

Another important step is simply to recognize that judgments based on intuition, prior experiences, observations, and anecdotal information occur widely in the field; but that these unavoidable components of the assessment process are rarely described as "informed clinical opinion." We believe that they are best described as *informed opinion* since such knowledgeable judgments are regularly used by parents and professionals alike. Clinical judgment can be simple, is often done unknowingly, and can be strengthened by providing a structure to the process that promotes functional and reliable outcomes and sound, comprehensive decisions.

Implications for Research

Notwithstanding the regulatory latitude for using clinical judgment in early intervention, basic research is still needed to examine: (1) the frequency with which

informed opinion and judgment is used for eligibility determination, (2) the procedures used to facilitate clinical judgment, and (3) the utility and accuracy of informed opinions and judgments when used for eligibility determination.

A comparison of outcome decisions for children with similar needs in states where informed opinion is used would certainly be of interest. In addition, studies are needed that examine the reliability, validity, and efficacy of existing judgment-based tools that can be used to facilitate this process. It is surprising that no studies of clinical judgment and its outcomes have been planned and conducted using state and county databases, particularly in light of the Part C allowance for the use of informed opinion (See Mott & Dunst, 2006, for an exception). This is one area where researchers and policy makers should conduct more rigorous research into documenting the utility and efficacy of clinical judgment.

Finally we recommend that the following operational definition of informed opinion be adopted for use in early intervention: Clinical judgment, or informed opinion, refers to the knowledgeable perceptions (i.e., impressions, intuitions, recollections, observations) of caregivers and professionals about the elusive and subtle capabilities of children in different settings that must be defined and quantified so that individuals or teams are able to reach accurate decisions about eligibility for early intervention.

Further, we advocate that the phrase, *informed opinion*, be adopted instead of clinical judgment or informed clinical opinion as the most communicable terminology to refer to the use of this process in early intervention. Since both parents and professionals can and do render decisions or judgments based on vital impressions, perceptions, intuitions, recollections, and observations, their decisions are not clinical in nature; rather, they are informed opinions based on personal life factors which are in and of themselves important in the decision-making process in early intervention.

A companion *Endpoints* (Vol. 2, Number 3) includes a summary of the findings presented in this *Cornerstones*. This nontechnical research summary also includes a description of those characteristics we consider the essential elements of clinical opinion when used for eligibility determination.

References

Egisdoittir, S., White, M., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., & Rush, J. D. (in press). The Meta-Analysis of Clinical Judgment Project: Fifty-six years of research on clinical versus mechanical prediction. *The Counseling Psy-*

- chologist (TCP) 34*: 341-382
- Bagnato, S. J. (1984). Team congruence in developmental diagnosis and intervention: Comparing clinical judgment and child performance measures. *School Psychology Review, 13*, 7-16.
- Bagnato, S. J., & Neisworth, J. T. (1985). Assessing young handicapped children: Clinical judgment versus developmental performance scales. *International Journal of Partial Hospitalization, 3*, 13-21.
- Bagnato, S. J., & Neisworth, J. T. (1990). *System to Plan Early Childhood Services (SPECS): Administration manual*. Circle Pines, MN: American Guidance Service.
- Bagnato, S. J., Neisworth, J. T., & McClosky, G. (1994). *System to Plan Early Childhood Services (SPECS)*. Circle Pines, MN: American Guidance Service.
- Bagnato, S. J., & Neisworth, J. T. (1999). Normative detection of early regulatory disorders and autism: Empirical confirmation of DC:0-3. *Infants and Young Children, 12*(2), 98-106.
- Bailey, D. B., Buysee, V., Simeonsson, R. J., Smith, T., & Keyes, L. (1995). Individual and team consensus ratings of child functioning. *Developmental Medicine and Child Neurology, 37*, 246-259.
- Bailey, D. B., Jr., Simeonsson, R. J., Buysse, V., & Smith, T. (1993). Reliability of an index of child characteristics. *Developmental Medicine and Child Neurology, 35*, 806-815.
- Bierman, K. L., Nix, R. L., Maples, J. J., & Murphy, S. A. (2006). Examining clinical judgment in an adaptive intervention design: The fast track program. *Journal of Consulting and Clinical Psychology, 74*, 468-481.
- Blacher-Dixon, J., & Simeonsson, R. J. (1981). Consistency and correspondence of mothers' and teachers' assessments of young handicapped children. *Journal of the Division for Early Childhood, 3*, 64-71.
- Casey, B. M., McIntyre, D. D., & Leveno, K. J. (2001). The continuing value of the Apgar score for the assessment of newborn infants. *New England Journal of Medicine, 344*, 467-471.
- Danaher, J., Shackelford, J., & Harbin, G. (2004). Revisiting a comparison of eligibility policies for infant/toddler programs and preschool special education programs. *Topics in Early Childhood Special Education, 24*, 59-67.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.
- Dunst, C. J., & Trivette, C. M. (2004). Toward a categorization scheme of child find, referral, early identification and eligibility determination practices. *Tracelines, 1*(2), 1-18. Available at http://www.tracecenter.info/tracelines/tracelines_vol1_no2.pdf
- Dunst, C. J., Trivette, C. M., Appl, D. J., & Bagnato, S. J. (2004). Framework for investigating child find, referral, early identification, and eligibility determination practices. *Tracelines, 1*(1), 1-11. Available at http://www.tracecenter.info/tracelines/tracelines_vol1_no1.pdf
- Fleischer, K. H., Belgredan, J. H., Bagnato, S. J., & Ogonosky, A. B. (1990). An overview of judgment-based assessment. *Topics in Early Childhood Special Education, 10*(3), 13-23.
- Garb, H. N. (1994). Toward a second generation of statistical prediction rules in psychodiagnosis and personality assessment. *Computers in Human Behavior, 10*, 377-394.
- Glascoc, F. P. (1991). Can clinical judgment detect children with speech-language problems? *Pediatrics, 87*, 317-322.
- Gradel, K., Thompson, M. S., & Sheehan, R. (1981). Parental and professional agreement in early childhood assessment. *Topics in Early Childhood Special Education, 1*(2), 31-39.
- Gresham, F. M., Reschly, D. J., & Carey, M. P. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. *School Psychology Review, 16*, 543-553.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.
- Hayes, A. (1990). The context and future of judgment-based assessment. *Topics in Early Childhood Special Education, 10*(3), 1-12.
- Hemmeter, M. L., Joseph, G. E., Smith, B. J., & Sandall, S. (2001). *DEC recommended practices: Program assessment: Improving practices for young children with special needs and their families*. Longmont, CO: Sopris West.
- Henderson, L. W., & Meisels, S. J. (1994). Parental involvement in the developmental screening of their young children: A multiple-source perspective. *Journal of Early Intervention, 18*, 141-154.
- Holt, R. R. (1970). Yet another look at clinical and statistical prediction: Or, is clinical psychology worthwhile? *American Psychologist, 25*, 337-349.
- Hunt, F. M., Johnson, C. R., Owen, G., Ormerod, A. J., & Babbitt, R. L. (1990). Early intervention for severe behavior problems: The use of judgment-based assessment procedures. *Topics in Early Childhood Special Education, 10*(3), 111-121.
- Individuals with Disabilities Education Act, 20 U.S.C. § 1432(5), 1435(a)(1) (1997).
- Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 *et seq* (2004).

- Kochanek, T., Kabacoff, R., & Lipsitt, L. (1990). Early identification of developmentally disabled and at-risk preschool children. *Exceptional Children, 56*, 528-538.
- LeLaurin, K. (1990). Judgment-based assessment: Making the implicit explicit. *Topics in Early Childhood Special Education, 10*(3), 96-110.
- LeVan, R. R. (1990). Clinical sampling in the assessment of young children with handicaps: Shopping for skills. *Topics in Early Childhood Special Education, 10*(3), 80-95.
- Marchese, M. C. (1992). Clinical versus actuarial prediction: A review of the literature. *Perceptual and Motor Skills, 75*, 583-594.
- McCloskey, G. (1990). Selecting and using early childhood rating scales. *Topics in Early Childhood Special Education, 10*(3), 39-64.
- Meehl, P. E. (1954). *Clinical versus statistical predictions: A theoretical analysis and review of the evidence*. Minneapolis, MN: University of Minnesota Press.
- Miltenberger, R. G. (1990). Assessment of treatment acceptability: A review of the literature. *Topics in Early Childhood Special Education, 10*(3), 24-38.
- Mott, D. W., & Dunst, C. J. (2006). Use of presumptive eligibility for enrolling children in Part C early intervention. *Journal of Early Intervention, 29*, 22-31.
- Records, N., & Tomblin, B. (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech and Hearing Research, 37*, 144-156.
- Sampers, J., Cooley, G., Cornelius, K., & Shook, L. (1996). Utilizing clinical judgment in the early identification of premature infants with motor difficulties. *Infant-Toddler Intervention: The Transdisciplinary Journal, 6*, 117-124.
- Sandall, S., McLean, M. E., & Smith, B. J. (Eds.). (2000). *DEC recommended practices in early intervention/early childhood special education*. Longmont, CO: Sopris West.
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178-200.
- Schalock, R. L., & Luckasson, R. (2005). *Clinical judgment*. Washington, DC: American Association on Mental Retardation.
- Sexton, D., Thompson, B., Perez, J., & Rheams, T. (1990). Maternal versus professional estimates of developmental status for young children with handicaps: An ecological approach. *Topics in Early Childhood Special Education, 10*(3), 80-95.
- Shackelford, J. (2002, May). Informed clinical opinion. *NECTAC Notes*, (No. 10), 1-4. Chapel Hill: National Early Childhood Technical Assistance Center, The University of North Carolina.
- Short, R. J., & Simeonsson, R. J. (1990). Stereotypical behaviors and handicapping conditions in infants and children. *Topics in Early Childhood Special Education, 10*(3), 122-130.
- Simeonsson, R. J., Bailey, D. B., Smith, T., & Buysee, V. (1995). Young children with disabilities: Functional assessment by teachers. *Journal of Developmental and Physical Disabilities, 7*, 267-284.
- Simeonsson, R. J., Huntington, G. S., Short, R. J., & Ware, W. B. (1982). The Carolina record of individual behavior: Characteristics of handicapped infants and children. *Topics in Early Childhood Special Education, 2*(2), 43-55.
- Sines, J. O. (1970). Actuarial versus clinical prediction in psychopathology. *British Journal of Psychiatry, 116*, 129-144.
- Suen, H., Lu, C.-H., Neisworth, J. T., & Bagnato, S. J. (1993). Measurement of teams decision-making through generalizability theory. *Journal of Psychoeducational Assessment, 11*, 120-132.
- Suen, H. K., Logan, C. R., Neisworth, J. T., & Bagnato, S. J. (1995). Professional congruence: Is it necessary? *Council for Exceptional Children, 19*, 243-252.
- Westen, D., & Weinberger, J. (2004). When clinical description becomes statistical prediction. *American Psychologist, 59*, 595-613.
- Wiggins, J. S. (1981). Clinical and statistical prediction: Where are we and where do we go from here? *Clinical Psychology Review, 1*, 3-18.

Authors

Stephen J. Bagnato, Ed.D., NCSP, is Professor of Pediatrics and Psychology; Co-Investigator and Pennsylvania Satellite Director for Tracking, Referral, and Assessment Center for Excellence (TRACE); and Director of Early Childhood Partnerships, Children's Hospital of Pittsburgh and The UCLID Center at the University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania (steve.bagnato@chp.edu; www.uclid.org). Janell Smith Jones, Ph.D., has been a Research Psychologist for the TRACE Children's Hospital of Pittsburgh, Pennsylvania (janell.smithjones@chp.edu). Margaret Matesa, M.Ed., is Coordinator of the Pennsylvania TRACE Satellite, Children's Hospital of Pittsburgh, Pennsylvania (margaret.matesa@chp.edu). Eileen McKeating-Esterle, M.S., is a Research Associate for the TRACE Center, Children's Hospital of Pittsburgh, Pennsylvania (eileen.mckeating-esterle@chp.edu).

Table 1
Background Characteristics of the Study Participants

Study	Sample Size	Mean Age (Months)	Age Range (Months)	Child Diagnosis
Bagnato (1984)	54	31	6–53	Multiple disabilities
Bagnato & Neisworth (1985)	58	34	16–62	Developmental disorders: developmental delay; mental retardation, neurological, communication, visual disabilities and behavior disorders
Bagnato, Neisworth & McClosky (1994)	1300	NR	24–71	Neuromotor and hearing impairments, communication and behavior/emotional disorders, developmental delay, mental retardation, autism/pervasive developmental disorders, and typical development
Bagnato & Neisworth (1999)	85	NR	15–45	Autism, fragile X syndrome, and general developmental delay
Bailey, Buysse, Simeonsson, Smith & Keyes (1995)	129	63.0	1–204	Vision and hearing loss, disabilities in social skills, behavior and cognitive functioning, limb impairment and communication disabilities
Bailey, Simeonsson, Buysse & Smith (1993)	254	53	14–104	Children had been previously identified as having some disability and were eligible for early intervention services
Bierman, Nix, Maples & Murphy (2006)	891 445 for sample 446 for control	79	72–108	Aggressive, disruptive children
Blacher-Dixon & Simeonsson (1981)	52	38	9–75	Mental retardation with various accompanying developmental disabilities
Casey, McIntire & Leveno (2001)	151,891	1 day	1 day	Pre- or full-term infant (*follow-up of survival monitored for first 28 days of life)
Glascoe (1991)	157	35	6–77	Language delay or typical development
Gradel, Thompson & Sheehan (1981)	60	NR	3–73	Developmental delay
Gresham, Reschly & Carey (1987)	200	111	90–138	Learning disability or typical development
Henderson & Meisels (1994)	90	62	47–71	Learning disability or typical development
Kochanek, Kabacoff & Lipsitt (1990)	536	NR	0–95	Behavior disorder, developmentally delayed, learning disability, or typical development (*follow-up with school records when children were between ages 14-20 years)
Records & Tomblin (1994)	92	NA (hypothetical case studies)	48–119	Language impairment or typical development
Sampers, Cooley, Cornelius & Shook (1996)	34	4	4	Developmental delay or typical development (*confirmation of developmental status at 24 months corrected age)
Simeonsson, Bailey, Smith & Buysse (1995)	108	47	36–69	Speech/language impaired, mentally handicapped, behaviorally/emotionally handicapped, multiply handicapped, autism, orthopedic impairment, learning disabled, other health impairment
Simeonsson, Huntington, Short & Ware (1982)	360	36	3–89	Mental retardation, hearing and visual loss, orthopedic impairment and speech and emotional disorders, singly or in combination
Suen, Lu, Neisworth & Bagnato (1993)	467	NR	45–60	Developmental delay or typical development
Suen, Logan, Neisworth & Bagnato (1995)	467	NR	45–60	Developmental delay or typical development

NR = Not reported.
 NA = Not applicable.

Table 2
Study Focus and Practice Characteristics

Study	Study Focus	Practice Characteristics				
		Defined Characteristics	Structured Opinions	Multiple Settings or Individuals	Multisource Consensus Decisions	Training to Structure Opinions
Bagnato (1984)	Parent-professional congruence	X	X	X	X	X
Bagnato & Neisworth (1985)	Parent-professional congruence	X	X	X		
Bagnato, Neisworth & McClosky (1994)	Judgment-based scales and formats	X	X	X		X
Bagnato & Neisworth (1999)	Judgment-based scales and formats	X	X	X		X
Bailey, Buysse, Simeonsson, Smith & Keyes (1995)	Professional congruence	X	X	X	X	
Bailey, Simeonsson, Buysse & Smith (1993)	Judgment-based scales and formats	X	X	X	X	X
Bierman, Nix, Maples & Murphy (2006)	Parent-professional congruence	X	X	X	X	X
Blacher-Dixon & Simeonsson (1981)	Parent-professional congruence	X	X	X		
Casey, McIntire & Leveno (2001)	Judgment-based scales and formats	X	X			X
Glascoe (1991)	Early detection and classification	X			X	
Gradel, Thompson & Sheehan (1981)	Parent-professional congruence	X	X	X		
Gresham, Reschly & Carey (1987)	Judgment-based scales and formats	X	X	X	X	
Henderson & Meisels (1994)	Early detection and classification	X	X	X	X	
Kochanek, Kabacoff & Lipsitt (1990)	Early detection and classification	X		X	X	
Records & Tomblin (1994)	Early detection and classification	X	X	X	X	
Sampers, Cooley, Cornelius & Shook (1996)	Early detection and classification	X	X			X
Simeonsson, Bailey, Smith & Buysse (1995)	Judgment-based scales and formats	X	X		X	
Simeonsson, Huntington, Short & Ware (1982)	Judgment-based scales and formats	X				
Suen, Lu, Neisworth & Bagnato (1993)	Judgment-based scales and formats	X	X	X	X	X
Suen, Logan, Neisworth & Bagnato (1995)	Judgment-based scales and formats	X	X	X	X	X
Percentage of studies		100	85	75	60	45

Table 3
Results and Major Findings Reported in Each of the Studies

Study	Study Type	Results	Major Findings
Bagnato (1984)	Case	<i>Congruence</i> Moderate to strong agreement between clinical judgment measures and other child performance measures ($r = .65, .66, \& .89$).	1. The use of carefully selected child performance and clinical judgment scales facilitate consistent, comprehensive, and accurate child assessments by teams by structuring and quantifying their subjective impressions of children's skills.
Bagnato & Neisworth (1985)	Case	<i>Congruence</i> Strong correlation (.81–83) between clinical judgment measures and norm- and curriculum-based measures rated by professionals. Low to moderate correlation (.34–69) between parent-rated clinical judgment measure and norm-based, curriculum-based, and clinical judgment measures rated by professionals.	1. Significant level of agreement between estimates of children's development on measures completed by teachers, SPL therapists and psychologists supported the practice of combining norm-referenced, criterion-referenced and clinical judgment scales within diagnostic batteries for young disabled children. 2. Less strong level of agreement between Parental and professional estimates of children's development. However, parental judgments of subtle changes in children's behavior (e.g. attention, activity level) were consistent with professional perceptions.
Bagnato, Neisworth & McClosky (1994)	Case	<i>Sensitivity</i> 1. Accurate classification of disability status: The D-SPECS correctly classified 84% of children with disabilities. <i>Specificity</i> 2. Accurate classification of nondisability status: The D-SPECS correctly classified 85% of children that did not have a disability.	1. Clinical judgment tool (D-SPECS) demonstrates a high degree of accuracy in discriminating between children with and without problems across multiple disability categories.
Bagnato & Neisworth (1999)	Case	<i>Accuracy</i> TABS correctly classified 83% of all children in sample (i.e., not at risk, at risk, and disabled). <i>Sensitivity</i> 72% of children correctly classified as needing early intervention services.	1. Normative data collected by the TABS appears to be congruent with the DC: 0-3 classification system and provides a means for parents and professionals to document their clinical opinions of self-regulatory issues. 2. Early symptoms of certain diagnostic classifications, such as autism, include issues in self-regulation and temperament. Earlier identification through tools such as the Temperament and Atypical Behavioral Scale (TABS) may pave the way for earlier interventions.
Bailey, Buysse, Simeonsson, Smith & Keyes (1995)	Case	<i>Congruence</i> 1. Across professionals within a team there was a relatively high degree of independent agreement on the child's general level of functioning (60% exact agreement, 87% agreement within one point). 2. Severity of disability emerged as the child characteristic that most significantly effected agreement. Fifteen of the 19 possible ratings on the index were more likely to cause disagreement when applied to children whose functional limitations were generally considered more severe (consensus ratings > 3).	1. Team members had a high degree of agreement on independent ratings of abilities and limitations. More disagreement was associated with rating areas of behavior and communication, team experience and increased severity of disability. 2. Team consensus ratings were more likely to defer to the ratings of the expert in particular areas.

Table 3, continued

Study	Study Type	Results	Major Findings
Bailey, Simeonsson, Buysse & Smith (1993)	Case	<p><i>Congruence</i></p> <ol style="list-style-type: none"> 1. Exact agreement across all three comparison groups averaged 67.2% agreement; agreement within one point across the three comparison groups averaged 86.2%. 2. When intraclass correlation procedures were applied, teacher-specialist agreement (mean = .73) was somewhat higher than the reliability between parents and professionals (mean = .60). 3. Considerable variation in agreement was found across individual items, however it was consistent across comparison pairs. 	<ol style="list-style-type: none"> 1. Raters who vary considerably in expertise, discipline, and relationship to a child generally agree (rate congruently) on overall estimations of their abilities. 2. Caution should be exercised in using individual item scores rated by a single individual for research or diagnostic purposes other than providing a description of the subjects in the sample, or a group of children served. The low reliability of several individual items suggests that raters should be trained and agreement should be monitored.
Bierman, Nix, Maples & Murphy (2006)	Cohort	<ol style="list-style-type: none"> 1. 95% of the selected sample scored in the top 20% on both the parent and teacher screening measures. 2. 76% of these children had teacher ratings in the clinical range on the Externalizing scale of the Child Behavior Checklist. 3. Two types of clinical judgments (family coordinator's rating of parental functioning, and their global assessments of family need for home visiting) were significantly correlated. Stronger reliability and better concurrent and predictive validity emerged for the 1st. 4. Almost all family coordinators increased or decreased in home visit dose recommendations over the four time periods covered by the study. 	<ol style="list-style-type: none"> 1. Family coordinators appeared to be more reliable in their ratings of parental functioning than in their global assessments of need for home visits. 2. It was beneficial to use clinical judgment to tailor dose recommendations for home visiting. 3. Clinicians are relatively strong in providing sensitive descriptions of client functioning when completing standardized rating scales in specific domains linked empirically with the development or remediation of a problem.
Blacher-Dixon & Simeonsson (1981)	Cohort	<p><i>Congruence</i></p> <ol style="list-style-type: none"> 1. Overall, weak to moderate agreement ($r = .27-.83$) between mothers and teachers, all correlations significant at $p < .05$. 2. Maternal ratings ($r = .62-.85$) were consistent over time on 50% of items on clinical judgment measure (CRIB). 	<ol style="list-style-type: none"> 1. Significant degree of agreement between mothers and teachers ratings of young handicapped children's behavioral skills. 2. Mothers were consistent in their ratings of selective behaviors and temperament characteristics over time.
Casey, McIntire & Leveno (2001)	Cohort	<p><i>Accuracy</i></p> <p>Pre-term infants; incidence of neonatal death: .315 per 1000 for newborns with 5-minute APGAR scores of 0-3, but only 5 per 1000 for newborns with scores of > 7 (95% confidence interval). Neonatal death rate highest for 5-minute APGAR scores of ≤ 3 regardless of gestational age ($p = .002$).</p> <p>Full-term infants; incidence of neonatal death: .244 per 1000 for newborns with scores of 0-3 whereas 0.2 per 1000 for infants with scores of ≥ 7 (.001).</p>	<ol style="list-style-type: none"> 1. Survival in preterm and term infants increases as APGAR scores increase. 2. Five-minute scores are more useful in predicting neonatal deaths than one-minute APGAR scores.
Glascoe (1991)	Case	<p><i>Sensitivity</i></p> <ol style="list-style-type: none"> 1. Moderately accurate prediction of speech/language impairments based on parental concerns: 72% of children who failed developmental assessment had parents with concerns. <p><i>Specificity</i></p> <ol style="list-style-type: none"> 2. Accurate prediction of children without speech/language impairments: 83% of children who passed developmental assessment had parents that did not have concerns. 	<ol style="list-style-type: none"> 1. Clinical judgment, defined in this study as eliciting and responding to parents' speech/language concerns for their child, is a useful prescreening strategy. However, it requires confirmation by standardized screening tests as notably more parents had speech and language concerns than children who actually had speech/language deficits.

Table 3, continued

Study	Study Type	Results	Major Findings
Gadel, Thompson & Sheehan (1981)	Case	<p><i>Congruence</i></p> <ol style="list-style-type: none"> Moderately weak to strong correlation ($r = .42-.87$) between mothers of infants and teachers' ratings on the clinical judgment measure (Developmental Profile). Strong agreement between mothers of preschoolers and teachers' ratings ($r = .95-.98$). 	<ol style="list-style-type: none"> Substantial agreement between mothers' and professionals' ratings (teachers and diagnosticians) of the skills of young children with disabilities. Parent-professional rating agreement was greater for mothers of preschoolers than mothers of infants.
Gresham, Reschly & Carey (1987)	Case	<p><i>Sensitivity</i></p> <ol style="list-style-type: none"> Clinical judgment measure (TRAP) correctly classified 91% of children with learning disabilities. <p><i>Specificity</i></p> <ol style="list-style-type: none"> TRAP correctly classified 86% of children that did not have a disability. 100% of learning disabled students and 86% of students without disabilities were classified correctly when standardized and clinical judgment measures were used together. 	<ol style="list-style-type: none"> Teachers are highly accurate in their classifications of children as learning disabled or non-disabled. Teacher ratings on a clinical judgment tool identified learning disabled students at a slightly higher rate than standardized IQ and achievement measures. Children's performances on standardized intelligence and achievement measures were slightly more accurate in identifying children without disabilities.
Henderson & Meisels (1994)	Cohort	<p><i>Sensitivity</i></p> <ol style="list-style-type: none"> Accurate prediction of developmental delay based on examiner and parent information (clinical judgment tool) decreased classification accuracy from 100% to 83%. Higher rate of accuracy associated with use of examiner only information. <p><i>Specificity</i></p> <ol style="list-style-type: none"> Accurate prediction of nondisability status based on examiner and parent information increased accuracy from 83% to 94%. 	<ol style="list-style-type: none"> Combining parent and examiner information decreased the effectiveness of identifying children with delays that actually needed further evaluation. Combining parent and examiner assessment information results in more effective screening of children not at risk for developmental delays and decreased substantially the rate of over referrals based on examiner information alone.
Kochanek, Kabacoff & Lipsitt (1990)	Cohort	<p><i>Sensitivity</i></p> <p>Weak to moderate accuracy in predicting disability status (30%–74%) depending on variable under consideration. Family factors 52%, child performance factors 30%–74%.</p> <p><i>Specificity</i></p> <p>Moderate to strong accuracy in predicting non-disability status (68%–92%) depending on variable under consideration. Family factors 62%, child performance factors 82%–92%.</p>	<ol style="list-style-type: none"> Risk factors differ in their impact on prediction of children's disability status at different points in time. Maternal education is a more accurate predictor of adolescent disability status than children's skill levels or behaviors between birth and 3 years of age. Children's skill levels and behaviors between ages 4 and 7 years are more accurate predictors of disabilities in adolescence than family factors.
Records & Tomblin (1994)	Case	<p><i>Congruence</i></p> <p>Considerable congruence between raters' diagnostic decisions of hypothetical cases.</p> <ol style="list-style-type: none"> Interrater reliability on 15 repeated cases was moderate (phi coefficient of .68; range of agreement 73%–100%). Language comprehension, nonverbal IQ, and language production assessment scores most heavily influenced diagnostic decisions. Better than chance agreement ($K = .14, p < .001$) between rater diagnoses. Language impairment cases judged as low normal or borderline in severity were associated with lower agreement among raters. 	<ol style="list-style-type: none"> It is possible to statistically model the overall diagnostic decision rule used by clinicians. Clinicians integrate multiple sources of information and attach greater importance to information from specific sources. Better than chance agreement between rater diagnoses. Raters' confidence in their diagnostic decisions were lower and more variable when judging borderline scores/skill levels.

Table 3, continued

Study	StudyType	Results	Major Findings
Sampers, Cooley, Cornelius & Shook (1996)	Cohort	<p><i>Sensitivity</i> 80% of children with high risk MAI scores at 4 months corrected age had abnormal Bayley scores at 24 months corrected age.</p> <p><i>Specificity</i> 79% of children with low risk MAI scores at 4 months corrected age had WNL Bayley scores 24 months corrected age.</p>	<ol style="list-style-type: none"> 1. Movement Assessment Inventory (MAI) is an effective clinical judgment tool for identifying preterm infants who were at significant risk for poor motor development and would benefit from early intervention services. 2. There was a significant relationship between early motor skill scores and later motor skill delays.
Simeonsson, Bailey, Smith & Buysse (1995)	Case	<ol style="list-style-type: none"> 1. The ABILITIES Index is a clinical judgment tool that provided a profile of abilities and disabilities unique to individual children as opposed to groups of children (e.g., autistic, learning disabled). 2. A sampling of skill/deficit profiles for individual children confirmed primary deficit areas, however, children were often presented with disabilities across several other deficit areas. 3. The tool found considerable variability in children's skills and needs within individual disability categories. 	<ol style="list-style-type: none"> 1. Case data provided visual evidence of the overlap of inter- and intra-individual variability among exceptional children. 2. This study provided preliminary evidence for the feasibility of documenting children's abilities, and have shown that documentation can be accomplished with satisfactory reliability among several raters, especially parents.
Simeonsson, Huntington, Short & Ware (1982)	Case	<p><i>Accuracy</i></p> <ol style="list-style-type: none"> 1. Mean CRIB subscale A scores (developmental) for the total sample ranged from 4.8 to 7.2, with SD ranging from 2.1 to 2.9. Corresponding A subscale values for the subgroups revealed lower mean scores for children who were mentally retarded or who had more than one handicapped condition, whereas somewhat higher mean scores were found for children with auditory or orthopedic impairments or Down syndrome. 2. Mentally retarded and multihandicapped children had subscale B scores (behavioral) that differed from the optimal score on the average of 1.5 units, whereas for the other three subgroups average discrepancies ranged from .5 –.9 units. 	<ol style="list-style-type: none"> 1. The CRIB was found to be a sensitive measure of individual differences and behavioral characteristics in handicapped children functioning in the infancy and toddler range.
Suen, Lu, Neisworth & Bagnato (1993)	Generalizability	<p><i>Accuracy</i></p> <p>Increasing reliability ($r = .57-.93$) and decreasing standard errors of measurement (.11 & .13) occurred on the Developmental SPECS when the number of raters (2–4) increased from two to four people and their information was pooled. Moderate to strong reliabilities (.71–.93) were found for multiple raters.</p>	<ol style="list-style-type: none"> 1. The most appropriate assessment outcomes and decisions are based on pooling information from multiple raters as combining information provides rich and accurate child assessments. 2. Parental ratings of children's skills are integral parts of the assessment process. Insistence on parent/professional congruence rather than pooling information diminishes the validity of parental ratings.
Suen, Logan, Neisworth & Bagnato (1995)	Generalizability	<p><i>Accuracy</i></p> <ol style="list-style-type: none"> 1. Scores submitted to generalizability analysis via GENOVA. Reliability estimates were high (.95 & .93) and standard errors were low (.11 & .13). 	<ol style="list-style-type: none"> 1. Information from parents and professionals best represents the status of the child when we recognize that each source has relevant, real, but incomplete information. Optimally information is pooled to arrive at comprehensive snapshots of children's strengths and needs. 2. Requiring that parent's assessments match professionals' assessments diminishes their social and ecological validity. 3. The desirability of multisource assessment information is upheld because of richness and its technical adequacy.