



Center for Excellence in Early Childhood Assessment
Pennsylvania Satellite  The UCLID Center at the University of Pittsburgh

BRIDGES: Practice-Based Research Syntheses
TRACE Center for Excellence

The Evidence for Clinical Judgment in Early Intervention

**Stephen J. Bagnato
Margie Matesa
Janell Smith-Jones
Antonio Fevola**

What is Clinical Judgment?

Despite the interest in and policy mandate for clinical judgment or “informed clinical opinion” in the early intervention field, no clear definition has been posed to describe what it is. Moreover, few well-designed research studies have been published to document the best methods, technical adequacy, and results of clinical judgment. Perhaps the most widely cited source in the early intervention field on clinical judgment or informed clinical opinion is Shackelford (2002) and its earlier versions (Biro, Daulton, Szanton, & Garner, 1992) distributed by NECTAS. Despite the informative value of this document, the authors offered no definition of the construct of “informed clinical opinion”, itself. The definition which was posed focused instead on the methods (“qualitative and quantitative information”), content (“difficult- to-measure aspects of developmental status”), and purpose (to determine status and need for early intervention) for clinical judgment. As reported by Hayes (1990), one of the better definitions of clinical judgment is attributed to Goodnow (1988) referring to assessments by parents and other lay individuals: “Clinical judgment, defined as inference or evaluation derived from intuition and/or personal experience, is the basis of many daily routine assessments by parents and professionals...” (p.2).

The definition and evidence base for clinical judgment with infants, toddlers, and preschool children is, at best, undefined. Nevertheless, early interventionists and policymakers advocate the importance of informed clinical

opinion for two major reasons: (1) state and federal Part C regulations promote the flexible use of informed clinical opinion with difficult-to-evaluate infants and toddlers who may have developmental difficulties; and (2) parents and team members can integrate broader, qualitative information (without formal tests) to answer difficult questions about the status and service needs of young children. Part C regulations mandate informed clinical opinion as one of several techniques “to determine the existence of a condition that has a high probability of resulting in developmental delay” (p.). Clinical judgment is regarded widely as an important adjunct when other more formal measures fail to be appropriate or useful. Moreover, informed clinical opinion offers the necessary flexibility to integrate data from interviews, observations of natural play behaviors, and data from multiple lay caregivers and records to guide the determination of probable delay or disability and to establish the basis for early intervention services.

Based upon our review of the literature for this research synthesis, we offer the following operational definition for clinical judgment or informed clinical opinion which we regard as synonymous.

***Clinical judgment or informed clinical opinion** refers to the observations and perceptions (i.e., impressions, recollections, intuitions, beliefs, feelings, inferences) of parents, interdisciplinary team professionals, and other caregivers which are structured, integrated, and then used to reach decisions, individually and/or collectively, about a child’s functional and behavioral attributes; the characteristics of the child’s caregiving environments; and the transactions among both in order to fulfill early intervention purposes (e.g., early identification, eligibility determination, program planning, and progress monitoring).*

What is the Historical Context for Clinical Judgment?

Since the 1950’s clinical judgment has been a widely studied concept or phenomenon, especially in psychology and medicine. Meehl’s (1954) seminal review comparing clinical and statistical prediction encouraged a flurry of studies of the phenomenon. However, the majority of the research was conducted in university-based laboratory settings and had very few implications for practical applications. A recent meta-analysis by Grove and colleagues (2000) examined studies which compared clinical versus mechanical prediction (statistical and actuarial) techniques in the psychological and medical literature. Recent research has demonstrated the greater accuracy of actuarial methods compared to clinical judgment methods (Dawes, Faust, & Meehl, 1989). While not directly germane to early intervention, the meta-analysis may have some application to alternative early detection strategies such as presumptive eligibility.

Nevertheless, clinical judgment has intuitive appeal and persists as an evaluation methodology in many health and human service fields. The nursing literature details the rationale for, teaching of, and application of clinical judgment in nurse-midwifery (Greener, 1988) and decision-making about life-threatening conditions (Benner & Tanner, 1987). The field of communication disorders has

produced research and position statements on the use of clinical judgments in a comprehensive decision-making process about the presence and type of speech/language disorder exhibited by individuals (Records & Weiss, 1990). In pediatric medicine, Glascoe and Dworkin (1993) have published research and position statements on the factors or “judgment heuristics” that influence physician’s abilities to discriminate typical from atypical child development.

Within special education and school psychology, researchers hold differing positions on the role, usefulness, and technical adequacy of clinical judgment in a broad assessment-based decision-making process. Gresham and colleagues (Gresham, Reschly & Carey, 1987) conducted research on the accuracy of teacher judgments in distinguishing between students of learning disabilities and those without learning difficulties. The results demonstrated that teachers judgments were as accurate as decisions based on the use of standardized tests of intelligence and achievement. Functional classification systems based on the use of clinical judgment in team decision-making processes have been proposed and examined, although only descriptively (Iscoe & Payne, 1972). Arguably, the most thorough study of research on assessment and decision-making in special education has been conducted by Ysseldyke and colleagues (Ysseldyke, Thurlow, Graden, Wesson, Algozzine, & Deno, 1983) through the University of Minnesota Learning Disabilities Research Institute. Specific to clinical judgment, the researchers determined that school psychologists and special education teachers were able to differentiate between students who were low-achieving or learning disabled with only 50% accuracy whereas introductory psychology students evidenced a 75% accuracy rate. Moreover, their research indicated that team decision-making processes about special education eligibility were flawed and based more often on non-performance data such as socio-economic status, family issues, and physical attractiveness.

What is the Evidence Base for the Use of Clinical Judgment in Early Intervention?

Few publications and still fewer high-quality research studies specific to early intervention have been published on clinical judgment and its associated variants such as informed clinical opinion and judgment-based evaluation.

Two of the most widely regarded and cited publications on clinical judgment are Neisworth (1990), and Meisels and Atkins-Burnett (2000). The only readily available source on clinical judgment in early intervention of its type is the special issue of *Topics in Early Childhood Special Education* devoted to “judgment-based assessment” (JBA) (Neisworth, 1990). Despite its unique focus and coverage, Neisworth (1990) contains nine position papers and literature reviews on clinical judgment and JBA and, thus, represents the predominant type of descriptive treatment of this topic in the literature. As the special issue states,

clinical judgment is widely practiced and valued in assessment for early intervention, especially for children with significant disabilities and to enable to parents to function as partners with professionals on interdisciplinary teams. Nevertheless, few research studies have been published that directly address the reliability, validity, and technical adequacy of clinical judgment or informed clinical opinion and its fulfillment of the purposes of early detection and eligibility determination. Rather, many studies in this area have focused on the adequacy of specific instruments such as rating scales.

Search Strategy, Sources, and Selection Criteria

Relevant position papers, literature reviews, and research studies, both published and unpublished, were identified by using the following search terms: *clinical judgment, informed clinical opinion, team decision-making, judgment-based assessment, subjective judgment, qualitative assessment*. Presuming that few well-designed studies of the phenomenon existed in the early intervention literature, we opted for an initial, broad search strategy which consisted of two levels. The first focused on any position papers, literature reviews, studies, and policy papers on the related topics across the interdisciplinary fields of psychology, medicine, allied health fields, developmental disabilities, and special education. This enabled us to gauge the breadth and depth of coverage for clinical judgment, in general. Then, we delimited our search by focusing on the early intervention literature, covering the birth to five year age range.

The primary information databases covered the following sources: Psychological Abstracts (PsychInfo), Social Sciences Citation Index, Education Resource Information Center (ERIC), Ovid, Medline, CINAHL, Health Source, Cochrane, and Medscape. Additionally, we conducted a secondary search on our own Endnote information database maintained by the primary author (SJB). In addition, we conducted a selective search of unpublished master's theses and doctoral dissertations on these related topics at the University of North Carolina, Penn State University, and the University of Pittsburgh. Finally, we conducted manual searches of the reference sections of books, journal articles, and chapters in early intervention to locate additional sources.

Again, concerned with the probable dearth of studies in this narrow area, our a priori criteria for the final selections were three broad emphases: clinical judgment as the topic, individuals with special needs of any age, and a particular focus in early childhood. Through this process, we sought to identify sources which offered an evidence base for clinical judgment that was specific to early intervention as well as research sources which held promise for application in early intervention.

Search Results Summary

During the course of our search, we collected 1164 sources related to clinical judgment, informed clinical opinion, and judgment-based decision-making. In our Endnotes, we have included 502 citations.

For this synthesis, we identified 19 studies from the interdisciplinary fields of medicine, psychology, speech/language pathology, special education, early childhood, and early intervention as our final selection of research with specific application or promise to form the evidence base in early intervention. Table 1 profiles the descriptive characteristics of the sample in each study, and the types of child delays/disabilities addressed. Table 2 profiles the type of research, measurement methodologies and analysis strategies, primary findings, and general conclusions. The search results can be divided into three broad categories of research studies: early detection and classification, parent-professional and team congruence, and judgment-based scales and formats.

Early Detection and Classification

In our broad spectrum review of the research literature, we were unable to identify any well-designed study which examined a large, representative sample and the application and accuracy of clinical judgment in the early detection and eligibility of infants and toddlers for early intervention services. Despite the Part C allowance for the use of informed clinical opinion, it seems surprising and concerning that no studies of clinical judgment and its outcomes have been planned and conducted using state and county databases.

Nevertheless, several early intervention studies have examined clinical judgment for early detection and classification of young children with delays/disabilities and suggest some promise about directions for future research on clinical judgment approaches (Kochanek, Kabacoff, & Lipsitt, 1990; Records & Tomblin, 1994; Glascoe, 1991; and Sampers, Cooley, Cornelius, & Shook, 1996).

The most sophisticated of these studies (Kochanek et al., 1990) used an index and control group design of a matched sample of 268 adolescents with and without disabilities who participated as infants in the National Collaborative Perinatal Project in Rhode Island. The purpose of the overall study was to examine whether combined child-centered data collected before age 7 and familial factors could predict adolescent disability using stepwise logistic regression analyses. Developmental data included both standardized test information and "...appraisal of developmental competency via clinical judgment...(p. 530). The results indicated that parental factors were better predictors than child-centered factors from birth to age 3. From ages 4-7, child-centered factors were better predictors. Overall, the researchers concluded that early identification models which focus exclusively on developmental delay

criteria or adverse medical events are not adequate to comprehensively identify early all children who are eventually judged to have disabilities.

Two studies addressed the early detection of young children with speech/language problems (Glascoe, 1991; Records & Tomblin, 1994). Both studies report research that purportedly supports the sensitivity, specificity, and reliability of parent and clinician judgments in identifying speech/language problems. Glascoe (1991) used the Parent's Evaluation of Developmental Status (PEDS), an experimental instrument with questionable psychometric qualities, to structure parent's judgments about early competencies. Of 157 parent-child dyads (ages 6 months to 77 months), specificity was 83%, sensitivity, 72% with false negatives and false positives of 28% and 17%, respectively. They concluded that the clinical judgments of parents can reasonably be substituted for regular use of standardized screening tests. Records and Tomblin (1994) used logistical regression analysis to examine "clinical decision-making" of 27 speech/language clinicians using 92 case profiles of children ages 4-10 years in order to detect the accuracy and level of agreement of diagnosing language impairment. Results showed significant inter-rater agreement among clinicians and a measurable decision-making pattern of using test information rather than less structured data to reach decisions.

Finally, clinical judgment was used to identify 34 preterm infants in a tertiary care NICU who were later determined to have motor difficulties. Specifically, the Movement Assessment of Infants (MAI) completed at 4 months of age through observational ratings by a physical therapist and a doctoral special education student were compared to motor assessments at 2 years of age on the Bayley Scales of Infant Development II (Sampers, et al., 1996). The results concluded that the MAI was effective for early identification and documentation of clinical judgment. However, the small and diverse sample, and threats to the internal validity of the procedure with the MAI make the results questionable.

Parent-Professional and Team Congruence

Despite the variability in research methods, the acceptable studies on the level of agreement among parents and professionals and within teams about the perceived developmental status of young children generate common conclusions. Parent's judgments are integral to accurate and representative assessments of the developmental status of young children. Teams which include the parents make accurate judgments about the needs of young children with developmental delays/disabilities. Clinical judgment measures and sources of information correspond well with more structured measures of developmental status.

Arguably, the two strongest studies in this area were conducted by Suen and colleagues (Suen, Lu, Neisworth & Bagnato, 1993; Suen, Logan, Neisworth,

& Bagnato, 1995). Both studies employed the same large sample of young children with (n=262) and without (n=205) developmental delays, independent clinical judgment ratings of parents and 30 professional teams on the System to Plan Early Childhood Services (SPECS) and a generalizability theory analysis strategy to determine the reliability and congruence of parent-professional team diagnostic assessments. Two results emerged from these related studies: (1) Teams decisions that use the collaborative judgments of four professionals and a parent produce the most reliable assessments of infants, toddlers, and preschoolers for high-stakes decisions; and (2) analyses of parent contributions to the assessment of young children should concentrate on the pooled assessment information of parents and professionals rather than conventional inter-rater agreement among parents and professionals which obscures the unique contribution of parent judgments about difficult-to-assess attributes. Parents and professionals are not interchangeable raters, but rather add unique contributions to the diagnostic assessment outcome. One of a representative group of studies (Henderson & Meisels, 1994) demonstrated that misclassifications of delay or risk for school failure in older preschoolers was reduced substantially when parent information was combined with an individually administered screening inventory on early learning. In turn, the predictive accuracy of the screening process was increased.

The remaining studies on parent-professional congruence, while reaching similar conclusions, are weak in the use of small samples (ns=54 and 58) and simple correlational designs and methods. Bagnato and colleagues (Bagnato, 1984; and Bagnato & Neisworth, 1985) studied the congruence among multiple measures of child status using parents and professional team members. The study examined the interrelations among conventional developmental scales (i.e., Gesell Developmental Schedules, Bayley Scales of Infant Development) and judgment-based rating scales (i.e., Preschool Attainment Record,, Perceptions of Developmental Status) with children showing significant disabilities. Results suggested that clinical judgment and developmental performance scales, and therefore, the assessments of parents and team members, were correlated and thus congruent in diagnosis of developmental delays. With similar limitations, Blacher-Dixon and Simeonsson (1985) studied the clinical judgment ratings of 52 mothers of children with developmental disabilities and their teachers on the Carolina Record of Individual Behavior (CRIB) and various temperament scales. The authors concluded that teachers and parents were congruent in their ratings and consistent across time in the assessment of difficulty-to-assess domains (e.g., social orientation, endurance, motivation). One of the most widely cited studies in the early intervention literature on parent-professional assessments is the research of Gradel and colleagues (Gradel, Thompson, & Sheehan, 1981). The study focused on 30 infants and 30 preschoolers and their mothers and preschool teachers. Using a combination of conventional and rating measures, the authors examined the congruence among parents and teachers and the patterns of agreement and disagreement in developmental assessments. The authors concluded that the

assessments of mothers and teachers were highly correlated, but more congruent for the older preschoolers, and that mothers systematically overestimated their children's performance.

Judgment-based Scales and Formats

Policymakers and researchers have neither operationally defined clinical judgment in the research literature, nor identified technically adequate methods to measure clinical judgment. Few judgment-based measures, relying on observations, judgments, recollections, and ratings, have demonstrated the necessary rigor for use with infants and toddlers exhibiting diverse developmental capabilities in early intervention. Nevertheless, several judgment-based instruments are unique, and some researchers have published research on the psychometric qualities which support the use of these methods in early intervention.

System to Plan Early Childhood Services (SPECS)

SPECS was designed to provide a structured format for gathering clinical judgment data among parents and professionals on an interdisciplinary team and applying the data for early detection and assessment (*Developmental Specs*), team decision-making (*Team Specs*), and program planning, service delivery and progress evaluation (*Program Specs*) about the needs of children from 2-6 years of age. The national field-validation research on SPECS is reported in two technical resource manuals (Bagnato, & Neisworth, 1990; Bagnato, Neisworth, & McCloskey, 1994).

Rating-rerating reliability studies on the *Developmental Specs* were conducted among early childhood educators and para-professional aides (n=163) with correlation coefficients ranging from .60 to .87 across the 19 developmental dimensions.

Concurrent validity studies were conducted between the *Developmental Specs* ratings across the clusters of communication, sensorimotor, physical, self-regulation, cognitive, and self-social, and 12 different formal and observational scales to document the comparability of the assessments and outcomes for children. The percentage of agreement among scales was moderate to high in all domains.

Discriminant function analyses were conducted on typical and atypical child data for various disability categories (e.g., delay, mental retardation, neuromotor impairment, communication disorder, hearing impairment, behavioral/emotional disorder, autism/PDD) and among various rater groups (e.g., parent, early childhood special educator teacher, speech/language pathologist, social worker, regular early childhood educator, para-professional aide, psychologist, occupational therapist). Analyses focused upon determining the classification accuracy of *Developmental Specs* ratings and SPECS

decisions about services for a mean of 594 children with typical development and a mean of 118 children with atypical development (range= 11-392). Analyses also focused on classification accuracy by rater group for over 200 team members. Overall mean correct classification by disability category was 84.8% (range= 75% for developmental delay to 96% for hearing impairment). Overall mean correct classification by rater group was 76.8% (range= 66% for psychologist to 90% for early childhood special educator).

Overall, the SPECS system has the necessary psychometric rigor for accurate and sensitive team decisions for eligibility determination, program planning, and progress monitoring.

The ABILITIES Index

The ABILITIES Index was designed to provide a non-categorical classification format for describing the functional capabilities of individuals with diverse developmental disabilities. ABILITIES uses a rating scale format (1-6) to classify a functional capability range from normal to profound difficulty. Cross-cultural research has been conducted to document the reliability of the ABILITIES index for use in making disability classification decisions (Bailey, Simeonsson, Buysse, & Smith, 1993; Simeonsson, Bailey, Smith, & Buysse, 1995). Studies involved 254 children, 213 parents, 133 teachers, and 135 interdisciplinary professionals.

Inter-rater agreement studies among 133 teachers and 135 specialists (total number of agreements/agreements + disagreements) showed that 86% of the ratings of one rater were within one point of those of the second rater. Correlations among raters were low to moderate with an average of .60 across raters. Lower ratings and levels of agreement were observed on less well-defined and complex categories (i.e., social skills, inappropriate behavior, intellectual functioning, communication, and health).

A study of the stability of ratings was conducted with 44 teachers on ratings conducted one month apart. Agreement within one point was documented in 91 % of the instances with kappas in the moderate range (.77).

Temperament and Atypical Behavior Scale (TABS)

The TABS is a judgment-based method to record observations and recollections about the presence or absence of atypical temperament and self-regulatory behaviors in children 11 to 71 months of age. TABS was designed specifically as a functional and non-categorical method to screen and determine

eligibility for young children for early intervention and support services who are at-risk for later developmental delays/disabilities. TABS is one of the few judgment-based instruments with national norms that has been field-validated for specific early intervention purposes.

National normative research on the TABS (Bagnato, Neisworth, Salvia, & Hunt, 1999; Bagnato & Neisworth, 1999) was conducted in 33 states with a pooled norm group of children with typical (n=621) and atypical development (n=212). Factor analytic studies on the full 55-item TABS Assessment resulted in four distinct and empirically derived factors with associated eigenvalues: detached (.52-.66); underreactive (.55-.66); hypersensitive/active (.55-.69); and dysregulated (.46-.61). These studies confirmed and validated the regulatory disorders conceptualization in the Diagnostic Classification System: 0-3 (Zero to Three, 1994).

The stability of TABS Assessment ratings were studied with a sample of 157 children over a 2-3 week period. All participated in early intervention programs, home- and center-based. Coefficients showed excellent stability ranging from .73 to .94.

Sensitivity and specificity studies on the TABS show strong support for the TABS screener. Of the 833 children in the pooled sample, 83% were correctly classified as normal or atypical by the screener. Of the 17% incorrectly classified, only 2.4% were false negatives and these were all children identified as at-risk not with disabilities. Approximately 14.5% of the children incorrectly classified were false positives; only 6% of these children showed disabilities. Therefore, a conservative estimate of the accuracy of the TABS screener is 72% correct classification.

Ages and Stages Questionnaire: Social-Emotional

The ASQ: SE (Squires, Bricker, & Twombly, 2002; Davis & Squires, 2002) is a norm-referenced observation-based rating scale designed to collect judgments about the social-emotional capabilities of infants, toddlers, and preschool children. Initial research on the ASQ:SE supports its reliability, validity, and overall technical adequacy as a screening and assessment instrument for use by parents and professionals in early intervention.

Normative data was based on over 3,000 rating scales and validity studies were based on over 1,000 children. Parent test-retest reliability data over a 1-3 week period was .94. Screening accuracy studies indicated overall sensitivity (correctly detecting delays or atypical development) at only 78% and overall specificity (correctly identifying typical development) at 94%. Concurrent validity comparisons between the ASQ:SE and other measures (TABS, BITSEA) on 90 children and mothers corroborate the normative studies on sensitivity and specificity.

Adaptive Behavior Assessment System II (ABAS)

While not described as a clinical judgment tool, the ABAS is, arguably, the most comprehensive and technically adequate judgment-based rating scale available to professionals. The ABAS is a norm-referenced, multi-domain and continuous measure of adaptive behavior competencies for individuals from birth to 89 years of age. The content of the ABAS is aligned with the DSM IV and AAMR systems. Raters include anyone, professional or layperson, who knows the individual best and can provide a representative appraisal of skills across multiple domains of functioning. For young children from birth to age five, the domains include communication, community use, functional pre-academics, home living, health and safety, leisure, self-care, self-direction, social, and motor. Caregiver raters respond through a 0-3 point rating scale: Is not able; never when needed, sometimes when needed, always when needed.

For children ages birth to five years, the standardization sample comprised 2,100 individuals representative of the US population. The ABAS II results in scores including a general adaptive composite (GAC) having a mean of 100 and standard deviation of 15 and skill area standard scores with a mean of 10 and standard deviation of 3, and critical values of the 90 and 95 percent confidence intervals. Reliability studies included inter-rater, test-retest, and internal consistency analyses. Perhaps the most unique feature of the ABAS II is the broad scope of validity studies comparing each atypical sample with a matched "typical" control group. The studies were conducted during development on individuals with at least 15 types of developmental disabilities and disorders (e.g., autism, learning disabilities, early developmental delays (n=126), Alzheimer's disease, neuromotor impairments). Concurrent validity studies were conducted with most of the major psychoeducational measures used by professionals. The validity study with children showing early developmental delays included a matched control group sample of typically developing children.

The APGAR

Perhaps the quintessential clinical judgment index is the APGAR to appraise and quantify the condition and prognosis for newborn infants. The APGAR (Apgar, 1952) is a 10-point judgment-based rating scale which classifies infant status in the areas of respiratory effort, muscle tone, reflex activity, heart rate, and color at 1 and 5 minutes following the birth of the child. Controversy has surrounded the technical adequacy of the APGAR and its predictive validity.

Casey and colleagues (Casey, McIntire, Leveno, 2001) published one of the largest studies to date in the *New England Journal of Medicine* to examine the prediction issue. The research team conducted a retrospective analysis of 151,891 live-born infants without malformations. The infants were delivered at 26 weeks gestation or later during the years 1988 to 1998. In addition, the researchers collected umbilical-artery blood pH values on 145,627 of the infants

and compared them to the APGAR to document which best predicted neonatal death early in the postnatal period. The researchers determined that "...the risk of neonatal death in term infants with 5-minute Apgar scores of 0 to 3 (relative risk, 1460; 95 percent confidence interval, 835 to 2555) was eight times the risk of term infants with umbilical-artery blood pH values of 7.0 or less (relative risk, 180, 95 percent confidence interval, 97 to 334)."

What Conclusions and Implications Can be Drawn About the Evidence base for Clinical Judgment in Early Intervention

Theory, policy, and clinical practice intersect and underscore the clear value and promise of clinical judgment or informed clinical opinion to address several purposes for evaluation in early intervention. Unfortunately, the existing research base is meager regarding the technical adequacy and utility of clinical judgment for evaluation. In general, the studies are not well-designed and are mostly simple correlation or concurrent validity studies with small sample sizes.

The major exceptions to this conclusion are the stepwise logistic regression analyses employed by Kochanek & Kabacoff (1990) in their early identification study of developmental delays/disabilities in a matched sample of 536 (index and control groups) children in the National Collaborative Perinatal Project (NCPNP); the discriminant function analyses for the System to Plan Early Childhood Services (Bagnato, Neisworth, & McCloskey, 1991; 1994) on over 1000 preschoolers with diverse developmental delays/disabilities; and the generalizability analyses conducted by Suen and colleagues (1985; 1986) to validate the indispensable contribution of parent assessments on teams and the validity of the "two heads are better than one" concept in interdisciplinary team assessments.

Within the early intervention literature, several conclusions can be drawn tentatively, about clinical judgment methods, but require more well-designed studies to validate the professional confidence placed in them and to ensure their valid use and promise within early intervention.

Assessment of Severe Disabilities

For children with severe developmental disabilities, clinical judgment methods in the form of rating scales, observation formats, and functional classification systems provide more detailed and sensitive information about status and change in the child's narrow and inconsistent behavioral repertoire and richer information for individualized intervention planning than conventional norm-referenced developmental and psychoeducational instruments.

Team Evaluations & Parents as Partners

Within interdisciplinary assessment teams for children with developmental delays/disabilities, each team member provides an important, individual contribution to the resultant team decision about child status. Team assessments provide a better “hit rate” about true child status than individual member assessments. Parents have indispensable and idiosyncratic information about child status and needs which when excluded from team assessments results in less reliable and valid appraisals.

Early Detection & Eligibility Determination

So few large-scale studies have been conducted in this area that no confident conclusions can be drawn about the use of clinical judgment for early detection of developmental delays/disabilities. However, based on the few studies in this area, we can conclude that early identification models which focus solely on developmental delay or adverse medical events from birth to three years of age are inadequate in fully identifying children who will be determined eventually to have measurable developmental disabilities. Multi-factorial risk models must be designed to synthesize data about child, family, and socio-economic contributors to the determination of “eligible” for early intervention.

Judgment-based Scales

Arguably the area with the most studies, relatively, shows that some assessment measures which structure and quantify observations and judgments about functional capabilities through ratings and classification formats demonstrate adequate reliability, validity, and utility to make diagnostic decisions and to fulfill early intervention purposes. Based on the available research literature, the following judgment-based evaluation measures can be used with confidence for screening, early detection, and assessment by early intervention professionals and parents. Instrument studies, especially on the SPECS, ABILITIES Index, TABS, ASQ:SE, and ABAS II demonstrate that it is possible to structure and quantify observations and resultant clinical judgments so that reliable, valid, and useful evaluations for early detection can occur.

The Concept of Judgment-based Assessment and Evaluation

While clinical judgment is the construct, judgment-based evaluation is the method to perform clinical judgment. *Judgment-based evaluation* structures and uses clinical judgments or informed clinical opinion to reach decisions. *Judgment-based evaluation* is a flexible, but systematic process through which individuals or teams gather, classify, integrate, and quantify their observations and perceptions across multiple environmental contexts to reach decisions about the child, the environment, and programmatic needs.

Judgment-based evaluation most often focuses on a child’s functional capabilities; ambiguous behavioral attributes; environmental characteristics that

may influence developmental progress (e.g., caregiver-child interactions, physical aspects of the home and preschool program settings); and changes in the child's competencies or behavior as a response to intervention.

Judgment-based evaluation helps to accomplish the following purposes:

1. To define eligibility for early intervention services when other more conventional means fail to be useful;
2. To detect capabilities that are inconsistent or low threshold in their expression;
3. To document small increments of change within an intervention program;
4. To unify and facilitate team decision-making about child characteristics and specific programmatic and intervention needs
5. To enable parents to be integral to the team assessment process

References

Bagnato, S. J. (1984). Team congruence in developmental diagnosis and intervention: Comparing clinical judgment and child performance measures. *School Psychology Review, 12*(1), 7-16.

Bagnato, S. J., & Neisworth, J. T. (1985). Assessing young handicapped children: Clinical judgment versus developmental performance scales. *Journal of Partial Hospitalizations, 3*, 13-21.

Bagnato, S.J., & Neisworth, J. T. (1990). *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.

Bagnato, S. J., Neisworth, J. T., & McCloskey, G. M. (1991). *System to Plan Early Childhood Services (SPECS)*. Circle Pines, MN: American Guidance Service.

Bagnato, S. J., Neisworth, J. T., & McCloskey, G. M. (1994). *System to Plan Early Childhood Services (SPECS)*. Circle Pines, MN: American Guidance Service.

Bagnato, S. J., Neisworth, J. T., Salvia, J., & Hunt, F. (1999). *Temperament and Atypical Behavior Scale (TABS)*. Baltimore: Paul H. Brookes Publishing Co., Inc.

Bailey, D. B., Simeonsson, R. J. Buysse, V., & Smith, T. (1993). Reliability of an index of child characteristics. *Developmental Medicine and Child Neurology*, 35, 806-815.

Benner, P., & Tanner, C. (1987). Clinical judgment: How expert nurses use intuition. *American Journal of Nursing*, (January), 23-31.

Biro, P., Daulton, D., & Szanton, E. (1991). Informed clinical opinion (NECTAC Notes No. 4). Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute, National Early Childhood Technical Assistance Center.

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.

Bricker, D. (Ed). (2002). *Assessment, Evaluation and Programming System for Infants and Children (2nd Ed.)*. Baltimore: Paul H. Brookes Publishing Co., Inc.

Casey, B. M., McIntire D. D., & Leveno, K. J. (2001). The continuing value or the apgar score for the assessment of newborn infants. *New England Journal of Medicine*, 344(7), 467-71.

Davis, M. S., & Squires, J. (2002, December). Comparing three recently developed social-emotional screening instruments. Paper presented at the meeting of the Division for Early Childhood (DEC), San Diego, CA.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*. 243, 1668-1674.

Glascoe, F. P. (1991). Can clinical judgment detect children with speech-language problems? *Pediatrics*, 87(3), 317-322.

Glascoe, F. P., & Dworkin, P. H. (1993). Obstacles to effective developmental surveillance: Errors in clinical reasoning. *Journal of Developmental and Behavioral Pediatrics*, 14(5), 344-349.

Goodnow, J. J. (1988). Parents' ideas, actions, and feelings: Models and methods from developmental and social psychology. *Child Development*, 59, 286-320.

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.

Greener, D. (1988). Clinical judgment in nurse-midwifery. A review of the research with implications for education. *Journal of Nurse-Midwifery*, 33(6), 261-268.

Gresham, F. M., Reschley, 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(1), 19-30.

Hayes, A (1990). The context and future of judgment-based assessment. *Topics in Early Childhood Special Education*, 10(3), 1-12.

Henderson, E. L., & Meisels, J. S. (1994). Parental involvement in the developmental screening of their young children: A multiple-source perspective. *Journal of Early Intervention*, 18(2), 141-154.

Hresko, W. P., Miquel, S. A., Sherbenou, R. J., & Burton, S. D. (1994). *Developmental Observation Checklist System*, Austin, TX: PRO-ED, Inc.

Iscoe, I. and S. Payne (1972). Development of a revised scale for the functional classification of exceptional children. In E. P. Trapp & P. Himelstein (Eds.), *Readings on exceptional children* (pp. 3-29). New York: Appleton-Century-Crofts.

Kochanek, T. T., Kabacoff, R. I., & Lipsitt, L. P. (1990). Early identification of developmentally disabled and at-risk preschool children. *Exceptional Children*, 56(6), 528-538.

Meehl, P. E. (1954). *Clinical versus statistical predictions: A theoretical analysis and review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Meisels, S. J., & Atkins-Burnett, S. (2000). The elements of early childhood assessment. In J. P. Shonkoff and S. Meisels (Eds.), *Handbook of early childhood intervention*. Cambridge, UK: Cambridge University Press.

Neisworth, J. T. (1990). Judgment-based assessment and social validity. *Topics in Early Childhood Special Education*, 10(3), _____

Records, N., & Tomblin, B. (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *American Speech-Language-Hearing Association*, 37, 144-156.

Records, N., & Weiss, A. L. (1990). Clinical judgment: An overview. *Journal of Childhood Communication Disorders*, 13(2), 153-165.

Sampers, J., Cooley, G., & Shook, L. (1996). Utilizing clinical judgment in the early identification of premature infants with motor difficulties. *Infant-Toddler Intervention*, 6(2), 117-124.

Shackelford, J. (2002). Informed clinical opinion. (NECTAC Notes No. 10). Chapel Hill, NC: The University of North Carolina FPG Child Development Institute National Early Childhood Technical Assistance Center.

Simeonsson, R., Bailey, D., Smith, T., & Buysee, V. (1995). Young children with disabilities: Functional assessment by teachers. *Journal of Developmental and Physical Disabilities*, 7(4), 267-284.

Squires, J., Bricker, D., & Twombly, E. (2002). *The Ages and Stages Questionnaire: Social Emotional*. Baltimore: Paul Brookes Publishing Co., Inc.

Suen, H. K., Logan, C. R., Neisworth, J. T., & Bagnato, S. J. (1995). Professional congruence: Is it necessary? *Council for Exceptional Children*, 19(3), 243-252.

Suen, H. K., Lu, C.H., Neisworth, J..T., & Bagnato, S. J. (1993). Measurement of team decision-making through generalizability theory. *Journal of Psychoeducational Assessment* 11, 120-132.

Ysseldyke, J., E., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S. (1983). Generalizations from five years of research on assessment and decision making: The University of Minnesota Institute, *EEQ*, 4(1), 75-93.

ZERO TO THREE: (1994). *Diagnostic Classification of Mental Health & Developmental Disorders of Infancy and Early Childhood (DC: 0-3)*. Washington, DC: ZERO TO THREE. National Center for Clinical Infant Programs.