The limitations of normative assessments of high-risk infants

Jack P. Shonkoff, MD
Assistant Professor of Pediatrics
Co-Director, Child Development Service
Department of Pediatrics
University of Massachusetts Medical School
Worcester, Massachusetts

The evaluation of human aptitude or achievement is rarely a neutral process. When it involves the assessment of competence in an infant, the impact of the experience on parents' evolving perceptions of their child can be substantial and long lasting. It is also important to recognize that developmental or behavioral testing and categorization is more often a cultural exercise than it is a clinical measurement of a biological property. Human assessments are never wholly objective or value-free, for fundamental questions regarding the boundary between normality and deviance have never been universally answered.

Where well-articulated criteria do exist, they are heavily influenced by prevailing social and political needs. Thus diagnostic categories of abnormal development may, in fact, not reflect intrinsic deficits in a child or family as much as they reveal a society's contemporary values (Sarason & Doris, 1979). Some critics have even suggested that many
developmental disabilities are cultural inventions that result in social handicaps that have a greater impact on individuals than the limitations associated with specific functional deficits (Gledman & Roth, 1980).

The demand for assessment of infants and toddlers in American society today is growing at a rapid pace. Multidisciplinary evaluation teams and organized early intervention programs for infants with identified problems or vulnerabilities have become increasingly available over the past decade (Beller, 1979), and the concept of early diagnosis and management for developmental-behavioral problems is gaining broad-based support (National Institute of Mental Health, 1979).

The impetus for this growing attention to very young children and their families comes from a multitude of sociopolitical factors, many of which have not been adequately studied in this context. Infants and toddlers are the subjects of greater scrutiny in this society than ever before. Although normative assessment techniques have been available for several decades, changing agendas in infant evaluation raise questions about their current appropriateness. An examination of the history of standardized tests provides a background for the understanding of the purposes and uses of these tests.

HISTORICAL ROOTS

Despite periodic rebellions, 20th-century America has had a sustained love affair with standardized testing of human abilities. As an alternative to selection methods based on social characteristics, such as family status or class, testing is viewed as a reflection of this society’s commitment to competition and meritocracy, as well as its general belief in the value of scientific measurement.

The use of normative evaluations, however, has never been well insulated from the existing political climate. In the early decades of the 20th century, for example, intelligence tests were often used in the United States to identify targets for discrimination or for systematic exclusion from society. Thus distinguished psychologists, such as Henry Goddard and Lewis Terman, assembled test data to provide scientific validation for popular efforts to legislate racist immigration restrictions and compulsory sterilization procedures for the mentally defective (Kamin, 1974).

In 1912 Goddard volunteered to test immigrants at Ellis Island in New York and found 83% of Jews, 80% of Hungarians, 79% of Italians, and 87% of Russians were “feeble-minded” (Goddard, 1913). In the first edition of the manual for the Stanford-Binet Scales, Terman (1916) wrote the following:

In the near future intelligence tests will bring tens of thousands of these high-grade defectives under the surveillance and protection of society. This will ultimately result in curtailing the reproduction of feeble-mindedness and in the elimination of an enormous amount of crime, pauperism, and industrial inefficiency. (p. 6)

From its harsh and exclusionary beginnings, the field of American psychometrics evolved to address a wide range of social purposes. The administration of standardized diagnostic instruments, for example, has served such diverse functions as attempting to understand why children fail to perform successfully in school (Anderson, 1982), comparatively ranking applicants for admission to professional preparation programs (Skager, 1982), providing data to facilitate career counseling (Friedman & Williams, 1982), and determining assignment of specific job responsibilities in industry and the military services (Wigdor & Garner, 1982). Although testing is now employed in social context that is more concerned with providing opportunities than with selective exclusion, battles continue over issues of test bias and interpretation of individual and group differences (Gordon & Terrell, 1981).

Infant and toddler testing has had its own interesting opamental a relative history. Dm expecting inborn inborn (Greven, 1974), industrial century, a concern of pediatric deemphas; children under diversion.

Emphasis compete relative America

promoted Holmes, I

Modern poetic me can families related to been most opamental a major for society (1)

The grown mental ac sequence of

Although employees can in the childhood be identified by instruments of middle-c normal
Emphasis on the developmental competence of preschool children is a relatively recent phenomenon in American history. During the colonial period, the Puritans focused on spiritual salvation through the imposition of rigid discipline to counteract the inborn sinful tendencies of young children (Greven, 1973; Wishy, 1968). In contrast, the industrialization and secularization of the 19th century, along with persistent high mortality rates of young children, prompted greater concern for physical health. In fact, many pediatric authorities in the late 1800s urged a deemphasis on educational stimulation of children under 5 years of age to prevent the diversion of vital forces from activities that promoted physical well-being (Griffith, 1895; Holmes, 1857).

Modern advances in preventive and therapeutic medicine have liberated most American families from these anxieties. As problems related to mortality and physical illness have been mastered, the "new morbidity" of developmental and behavioral concerns has become a major focus for an increasing segment of the society (Haggerty, Boglimann, & Pless, 1975). The growing demand for formal developmental assessment has been a natural consequence of that focus.

Although systematic infant scales were employed clinically by a New Orleans physician in the late 19th century (Charle, 1887; cited in Garwood, 1982), the concepts of early childhood assessment were initially popularized by Arnold Gesell (1925), whose first instruments were standardized on a white, middle-class sample that was designated as "normal" based on the absence of obvious clinical pathology. In their genesis as well as their use, the Gesell scales were based on a model that considered child development a reasonably stable, predictable, linear process in which the acquisition of new skills is incremental and determined essentially by maturational factors.

The construction of the scales was primarily focused on commonalities of performance and was not concerned with individual differences among children. Test items were selected according to their appeal to young children and the ease with which interobserver reliability for scoring could be achieved. They were not chosen because of their validity with respect to a specific conceptualization of the developmental process itself. Composite scores and statistical treatment of the performance data were avoided (Ball, 1977), and descriptivestatistics of the process of development were emphasized (Yang, 1979).

During the decades that followed their introduction, early childhood developmental scales were used extensively for preadoption evaluations (Knobloch & Pasamanick, 1974) and to determine candidacy for early institutionalization. The idea of assessment as a vehicle for early identification of deviance and intervention in the developmental process was nonexistent in those early years. Thus although infant testing was characterized as a measure of developmental status and not intelligence, it conformed to the dominant thrust of the early psychometric movement, which was to categorize people for discrimination rather than identify them as potential recipients of services. In an era when early intervention strategies were not available, the formal evaluation of infants led to the labeling of pathology and not its treatment.

Contemporary Social Issues

A careful inspection of current priorities for early childhood testing reveals important and, in many cases, radical differences from the
historical context of this testing. Although social and political factors continue to have significant impact on evolving concepts of assessment and labeling, these influences now emanate from a society that has undergone many critical changes since the days of Gesell and his followers. The tremendous increase in the numbers and categories of very young children whose developmental status is being formally evaluated has been greatly stimulated by these changes, including such factors as altered family relationships, shifting attitudes toward handicapped persons, and changes in focus for the fields of developmental psychology and pediatrics.

Most of the normative tests commonly used to evaluate infants today, including the Bayley Scales of Infant Development and the Denver Developmental Screening Test, are modifications of the original Gesell scales. However, there have been dramatic changes in the goals of the assessments. Evaluations are now designed to identify children in need of intervention services (Meier, 1973). Diagnosis is viewed as an indication for assistance rather than a condemnation to inevitable deviance (Gordon & Terrell, 1981).

A revolution in attitudes toward the education of disabled children and interest in compensatory preschool programs has created a growing demand for more highly discriminating assessment techniques for younger children. Expanded interest in the specific learning disabilities of school-aged youngsters has generated additional pressures to identify the precursors of such high-incidence disorders in the subtle dysfunctions and atypical styles of preschoolers (Denhoff, Hainsworth, & Hainsworth, 1972; Trehub, 1977).

It is the frustrated desire for more accurate, long-term prognostication that is the basis of a great deal of the current criticism of infant developmental scales. Some investigators have responded to this criticism with increasingly reductionistic tests in a quest to identify discrete attributes that correlate most highly with later school performance. Social interaction (Cameron, Livson, & Bayley, 1967), early verbal skills (Moore, 1967), habituation to stimuli (Lewis, 1971), and attentional abilities (Kagan, Lapidus, & Moore, 1978) are a few of the many areas of functioning that have been studied in this regard.

Although the search for sensitive and specific predictors of later development has been vigorous, it has been relatively fruitless. Despite moderate correlations between significant early delays and later deviation (Illingworth, 1961), no single infant test has been shown to have a high degree of long-term predictive validity for children without obvious handicaps (McCall, Hogarty, & Hurlburt, 1972), and the arguments continue between those who feel that the tests just need further refinement and those who feel that better predictability is impossible because of the nature of human development (McCall, 1982; Rutter, 1971; Thomas, 1967). An examination of the social context in which these battles are currently being waged provides a better understanding of the problems inherent in the continued reliance on traditional normative assessment techniques.

Changing status of infants and families

Demographic data clearly indicate that childbearing today is less automatic and much more carefully planned than it was in the past. The available options for voluntary control over conception and pregnancy are unprecedented in human history, and Americans are having fewer babies. The average young woman today expects to have two children, compared with four children three generations ago (Hayes, 1980). Whereas economic considerations formerly dictated the need for large families (especially in rural settings), they now produce pressure to limit family size among the middle class and working poor.

(McKissick, 1979)
at highly with interaction (67), early version to stimuliabilities (Kagan, a few of the at have been positive and positive, has been a very fruitful, between signification (Hillgast, test has been of long-term children without garty, & Hurvicts continue tests just need who feel that the because of extent (McCall, 67). An examination which these agents provide a blend inherent traditional normativist.

Families indicate that in the past, unitary control are impressive. Americans are average young two children, three generations economic the need for rural settings, with family size working poor


Moreover, increasing numbers of well-educated, upper-middle-class women are choosing to defer childbearing until their professional careers have been well established, and others are deciding to forgo parenthood entirely (Hayes, 1980). The end result, for all groups in society, is smaller families, with the likelihood of an increased investment in each individual child. The degree to which such investment serves a greater narcissistic function for contemporary parents than for parents in preceding generations requires further study.

The revolution in the status of women during the past decade has had major repercussions for their children as well as for family relationships. In 1948, 26% of married women with schoolchildren and 43% of those with preschool children were employed outside the home; in 1976 the numbers rose to 54% and 37%, respectively (Keniston & the Carnegie Council on Children, 1977). Recent studies have projected that a rapidly increasing percentage of the preschool children in this country will be members of "single-working-parent" or "two-career" families within the next 10 years (Kammerman & Kahn, 1981). Conservative estimates project that one-fourth of all children under 18 years of age will be living with one parent by the year 1990 (Hayes, 1980).

With the disappearance of the multigenerational, extended family, the demand for alternative child care arrangements beginning in earliest infancy is growing. Among the effects of these new social relationships on infants is that the infants are leaving the shelter of the family and entering society at a very young age. Even in the most nurturant and supportive day care environments, many young children are subject to the scrutiny of nonfamily members who may be influenced by growing demands for early identification and remediation of developmental difficulties. This early introduction to our competitive social system inevitably highlights differences in competence among young children and may very well contribute to increased referrals for diagnostic evaluations.

The parental role in this new social context is complex. It is ironic that the opportunity for greater investment in each individual child made possible by carefully planned, smaller families may be accompanied by an erosion of self-confidence regarding perceived parenting competence. Isolation from the traditional child-rearing wisdom of the extended family, a decrease in opportunities for learning about infant care previously available to older siblings in large families, and the assignment of major child care responsibilities to nonfamily members all contribute to a paucity of knowledge about normal child development among many parents. The exploding proliferation of child care manuals and the increasing demand for developmental assessments of young children with subtle problems clearly reflect the anxieties of many parents and their willingness to rely on experts in matters relating to their children's development (Keniston & the Carnegie Council on Children, 1977).

The interplay among the social factors described leaves the contemporary American infant in a vulnerable position. Parents’ narcissistic investment in their children, the dilution of their autonomy in child rearing, their diminished confidence in their own child development expertise, and the growing pressures on society to evaluate very young children in order to identify and remediate developmental problems are some of the many cultural influences whose impact has not been adequately studied.

As we enter the postindustrial era, concern for infants in the United States extends beyond the family to the society at large. The care and education of very young children have become issues with public policy implications,
The care and education of very young children have become issues with public policy implications, which are being debated in legislative and judicial bodies.

which are being debated in legislative and judicial bodies (Hayes, 1982). As these matters become increasingly oriented toward concerns regarding competence and the promotion of competitive advantage, the demands for formal developmental evaluations expand.

Changing attitudes toward disadvantaged groups

The civil rights movement of the 1960s established a national mandate for compensatory efforts for groups of individuals who are disadvantaged because of institutionalized discrimination. For children from the lower socioeconomic classes, this commitment was manifested in the emphasis on preschool educational enrichment programs, such as Head Start (Zigler & Valentine, 1979). Evaluative data on the impact of these programs have been extensive and somewhat inconsistent, but they do support the claim that certain preschool experiences can positively influence later school performance (Bronfenbrenner, 1974; Lazar, Hubbell, Murray, Rosehe, & Royce, 1977).

The revolution in societal attitudes toward handicapped persons shares many vital characteristics with the civil rights movement for racial minorities (Gliedman & Roth, 1980). As the horrors of institutionalization for mentally retarded individuals became widely publicized (Blatt & Kaplan, 1974) and as developmentally disabled persons have achieved greater visibility, legislative and judicial prohibitions have been mounted against a wide variety of discriminatory practices, including those in education, employment, and housing. Legal guarantees of appropriate educational opportunities have been accompanied by a noncategorical, nondiagnostic orientation that emphasizes the formulation of individualized educational plans based on each child’s unique profile of assets and deficits.

This philosophy has expanded the range of “special needs” from severe, low-incidence disabilities, such as significant retardation, to the subtle, high-incidence disabilities of specific learning disorders and dysfunctional learning styles (Levine, Brooks, & Shonkoff, 1980). Based on data supporting the critical impact of the early parent-child relationship and the demonstrated increased plasticity of the immature central nervous system, pressure has mounted for efforts to identify all developmentally vulnerable children as early as possible in order to prevent or help minimize later handicaps (Tjossem, 1976). The growing demand for infant assessment and high-risk follow-up programs are a natural part of that movement.

Changing theories and academic interest in the developmental psychology of infancy

Interest in child development as a subject for scientific study is a 20th-century phenomenon (Stern, 1975). Academic attention to infancy has virtually exploded during the past 20 years. Although the diversity of conceptualizations and theoretical biases has been substantial (from Hall and Gesell to Freud, Watson, and Piaget), most contemporary scholars view human development as a transactional process in which the influences of nature and nurture act in a dynamic system of mutuality. Developmental gains are viewed increasingly as qualitative changes that emerge over time, and not simply as discrete skills that appear according to a preprogrammed maturational timetable.
An appreciation of the importance of individual differences among children is another central theme of the literature of the past 2 decades. Whereas traditional models of child development assumed that any deviation from the standardized progression of acquired skills and abilities was abnormal, recent evidence from both clinical and cross-cultural studies suggests that there may be a variety of patterns of development in the first years of life that result in later outcomes with comparable competence.

Some of these variations, such as the precocious motor development of Ugandan infants, who are stimulated to sit and move earlier (Super, 1976), are culturally determined. Others, such as the relatively greater reliance on visual exploration in infants with less-well-developed fine motor skills (Kopp, 1974), reflect functional adaptations that are beneficial within the context of specific constitutional attributes in an individual child. These data support the concept that atypical patterns in individual children (within-group differences) and among different cultural groups (between-group differences) may be quite normal (Cole & Bruner, 1971; Lesser, Fifer, & Clark, 1965; LeVeen, 1980).

Adaptive patterns of development in biologically impaired children may also vary considerably. Examples of this phenomenon include recent speculations regarding alternative early pathways toward ultimate linguistic competence (Naremore, 1979), the demonstrated development of object permanence in children with congenital blindness (Fraftberg, 1977), and the clear mastery of cognitive concepts by children with limb deformities that preclude the usual manipulatory experiences through which most children learn (Decaric, 1965). In view of the newly recognized possibility of multiple adaptive variants in developmental profiles, as well as the established poor predictive validity of developmental testing of children under 3 years of age, it is becoming appreciated that an uncritical use of normative instruments can lead to inappropriate diagnoses of abnormality in young children.

**Changing nature of pediatric health care**

The changing nature of pediatric health services in this country has been well documented (Haggerty et al., 1975). Improved infant feeding practices, immunizations, and antibiotics have made more physician time available for dealing with issues regarding development and behavior. Moreover, dramatic advances in perinatal intensive care, pediatric surgery, and the management of chronic diseases have raised new challenges regarding the promotion of optimal developmental progress beginning in early infancy. The proliferation of medical center-based, multidisciplinary, developmental evaluation clinics attests to the increased desire for formal, medically oriented assessments in this area.

As pediatric involvement in the developmental assessment process has expanded, however, serious questions have been raised about the skills of pediatricians themselves. Numerous studies have found that the majority of practicing clinicians feel that their formal training was not oriented sufficiently toward the realities of a general pediatric practice (Dworkin, Shonkoff, Leviton, & Levine, 1979). Areas repeatedly stressed as requiring more educational emphasis include behavior problems, normal growth and development, mental retardation, emotional problems, and developmental disabilities. In a recently published monograph, the Task Force on Pediatric Education (1978) characterized the current biosocial content of many programs as inadequate, while simultaneously endorsing the need for pediatricians to become increasingly involved in the process of developmental assessment.
The current context

Important social trends in contemporary American society seem to have provided the fertile soil in which developmental assessment and intervention for vulnerable infants have grown. The changing status of women and families has increased the cultural visibility of very young children. Professional trends in pediatrics and developmental psychology and the dramatic shift in attitudes toward disabled persons have generated substantial investment in the early identification and potential amelioration of handicapping conditions. Opportunities for infant assessment have expanded, and the demand in this area appears to be continuing to increase.

A CRITICAL APPRAISAL OF NORMATIVE TESTS

Within the current social context, the process of normative evaluation and diagnosis is complex and potentially problematic. Much of the difficulty is related to new agendas for infant assessment and a failure to distinguish between the legitimate goal of promoting optimal progress in vulnerable youngsters and the potentially destructive act of assigning early labels to these with atypical developmental patterns.

The factors that contribute to the tension between the legacy of normative evaluation techniques and the complex tasks of contemporary infant assessment are readily apparent when one analyzes the properties of commonly used screening tools, such as the Denver Developmental Screening Test (DDST), and diagnostic instruments, such as the Bayley Scales of Infant Development (BSID).

Standardization

The BSID was standardized on a sample of 1,262 infants and children 2–30 months of age who were selected to reflect the composition of the population of the United States (Bayley, 1969). This normative group included children from urban and rural areas, and the accumulated data were controlled for differences in sex, race, and parent education. All 1,262 children were living at home and presumed to be normal. Children with a history of prematurity, or institutionalization and children living in bilingual homes were explicitly excluded from the sample.

The DDST was standardized on 1,036 children 2 weeks to 64 years of age who were selected to reflect the ethnic and occupational makeup of the population of Denver, Colorado (Frankenburg & Dodels, 1967). Adopted children, those who were premature, and those with known handicaps were eliminated from the sample. The use of these

An increasing body of data has been compiled that demonstrates the problem of applying a single set of normative values in assessing infants of different cultural groups.

developmental tests for diverse sociocultural populations and atypical groups of children presents important problems regarding interpretation.

An increasing body of data has been compiled that demonstrates the problem of applying a single set of normative values in assessing infants of different cultural groups. Large-scale screening efforts using the DDST in Japan and the British Isles, for example, revealed important discrepancies in performance, particularly in the areas of language and early motor skills.

Bryant, Davies, and Newcombe (1974) administered the DDST to a random sample of 668 Cardiff infants between 2 weeks and 12 months of age, excluding adopted and handi-
United States (Bayley, 1974) included children, and the rolled for different education. All home and pre-
with a history of and children were explicitly
nelled on 1,636 children who were ad occupational of Denver, Dodds, 1967) were premature, ages were eliminated.

has been the le set of of

se sociocultural ips of children regarding inter-
has been complen of applying in assessing groups. Large the DDST in for example, cies in perfor-
as of language
combe, 1974)andom sample of weeks and 12 fed and handi-
capped children, but including children born prematurely without evidence of obvious abnormalities. When compared with the Denver norms, the test results showed consistently slower gross motor development in conjunction with slightly accelerated development of language and personal-social skills. These findings were confirmed in a subsequent study of a random sample of 1,547 Cardiff children under 6 years of age, and advanced language was noted in Cardiff children up to 18 months of age (Bryant, Davies, & Newcombe, 1979).

Studies by Ueda (1978a, 1978b) in Okinawa (615 children) and Tokyo (1,171 children) revealed a pattern of discrepancies that differed from the normative performance of the original Denver sample. The implications of these data regarding the sensitivity and specificity of screening decisions for children from different cultures are obvious.

Cross-cultural studies using the Bayley and Gesell scales have demonstrated similar problems. Brazelton, Robey, and Coller (1969) documented a generally slower timetable for development of a broad range of psychomotor skills in the Zinacanteco Indians of Southern Mexico during the first 9 months of life. Observations of parent-infant interactions and child-rearing practices revealed consistent reinforcement from parents for infants' passivity and low motor activity and minimal positive feedback from adult care givers for infants' vocalizations, smiling, or motor development.

In a society in which conformity is highly respected and individual self-expression is not a goal, the observed slower rate of emergence of those skills that are clearly devalued by parents in the first year of life suggests that the pace of task mastery in early infancy is influenced by care-giving contingencies. Studies of precocious motor development in infants in East Africa, where early child care practices actively facilitate sitting and walking (Siper, 1976), support further the conclusion that dis-

regard for cultural differences in the use and interpretation of normative scales is highly problematic.

Differences among atypical groups of children within a common culture present even more complex challenges to the use of normative tools. Premature infants, children who sustained varying degrees of asphyxia at birth, and those with specific disabilities are some of the many young children who were excluded from the original standardization samples for the traditional normative scales and who make up a large proportion of those who are currently being evaluated with these tools. Although knowledge in this area is incomplete, available data support the assumption that the course of adaptive development within many of these groups may vary from the normative patterns described for healthy, full-term infants (Sameroff & Chandler, 1975).

The interpretation of atypical developmental profiles in a clinical setting is a critical issue. Variations in the normative patterns of skill acquisition for low-birth-weight infants in the first year of life are often different from those described for full-term infants (Kopp & Parmelee, 1979). The distinction between the initial signs of cerebral palsy and the transient dystonias often observed in many premature infants in the early months of life is often difficult to discern (Drillien, 1972). Neurodevelopmental abnormalities in an infant who is recovering from a perinatal hypoxic episode may reflect a temporary disorganization of the central nervous system that is compatible with recovery in later childhood (Broman, 1979). Of the youngsters whose atypical patterns of progress are devalued by normative testing, children with discrete disabilities are perhaps the most heavily penalized. When measured according to the standards of traditional normative scales, the performance of children with specific handicapping conditions is automatically categorized as defective. When applied and interpreted in the prescribed
manner, such criteria do not acknowledge the possible adaptive characteristics of some deviant patterns of development. As suggested by Gliedman and Roth (1980), some disabled children may acquire new skills according to a healthy logic of their own.

For children with isolated deficits in one or more areas of performance, seemingly pathological responses on normative tests may reflect alternative adaptive behaviors that can be understood only in the context of the child’s unique blend of assets and liabilities. Rational assessment of the developmental progress of atypical children (whether they be blind, deaf, spastic, or the vulnerable products of premature or traumatic deliveries) requires new data that reflect the ranges of performance for each population.

Predictive validity

The greatest hazards in the current use of normative testing for infants at risk for disabling conditions are issues regarding specific diagnoses and the formulation of long-term developmental prognosis. Although Gesell believed that developmental diagnosis implied prognosis, the predictive validity of infant testing has been shown to be generally poor (Lewis & McGurk, 1972; McCall et al., 1972). As noted by Bayley (1970): “It is now well established that test scores earned in the first year or two have relatively little predictive validity (in contrast to tests at school age or later), although they may have validity as measures of the children’s cognitive ability at the time” (p. 1174).

Since long-term prognosis is a central issue in the evaluation of high-risk infants, the limited predictability of normative scales raises critical questions regarding their use. One of the most frequent problems relates to the misleading significance of group statistics when applied to individual clinical cases. All group data conceal wide variations in individual developmental rates. Even when aggregate correlations are exceptionally high, the scores of individual children can differ by a full standard deviation in either direction. When correlations are low, as in the relationship between developmental testing in infancy and testing in later childhood, the formulation of long-term predictions is clearly unjustifiable. When such efforts are applied to atypical groups of children who were excluded from the original standardization samples, the results are especially hazardous.

The calculation of quantitative test scores in the form of developmental quotients or age levels is another aspect of normative testing that requires reconsideration. Concepts such as “performance at the 17-month level” and “6 months of progress in the past 8 months” invite misinterpretation and may create a sense of precision in developmental diagnosis that does not really exist. Developmental age levels are remnants of the old incremental model of development popularized by Binet and applied by Gesell to the clinical evaluation of infants and toddlers. In the context of contemporary theories of the developmental process, they are meaningless numbers that are determined by the simple summation of discrete successes on a series of tasks without regard to the child’s pattern of performance. Interpretations of the difference between developmental and chronological age levels are often arbitrary, and the distinctions between diagnoses such as “below average,” “slow,” and “developmentally delayed” are generally vague.

The concept of a corrected age for premature infants is another convention whose clinical application is potentially problematic. The adjustment of developmental expectations for preterm infants was proposed by Gesell and has been practiced by clinicians throughout the history of infant assessment (Knobloch & Pasamanick, 1974). The concept of corrected age, however, has never been statistically validated, and its use is particularly susceptible to distortion.

In infants, Morris mental deviant scores a analysis of standardized tests problems for across all of over rejected diagnosis were for prenatal

ALTERED NORMS

In view with the tion of for assessing efforts in strategies traditions development. Clinita employee. As the p continuous, the monitor. young child programs, paired phy 1979). Modernized instead of comprehensive and office.
In a stud...
When aggregate scores are used, the scores are cast in a manner by a full scale norm. When the relationship between the norms and the administration of the test is unjustifiable, the test is considered to be atypical. The conclusion derived from the new samples, the test scores in patients or age normative testing concepts such as "chronological level" and "6 months" invite a sense of diagnosis that does not fit the age levels are determined primarily by Binet and others. Evaluation of the test of loneliness within a normative process, they determine the discrete norm and regard to diagnosis. Interpretative developmental often arbitrarily diagnoses "good" and "developing" vague.

For prematurity, the clinical situation. The expectations for Gesell and throughout Knobloch & others, are essentially validated and incapable of distortion in assessing infants in the early months of life (Hunt & Rhodes, 1977; Parmelee & Schulte, 1970).

In a clinical investigation of 413 preterm infants between 6 and 26 months of age, Morrison (1980) compared BSID median mental developmental indices (MDI) with the standardization data and found significantly lower scores at 8, 10, 12, and 24 months and significantly higher scores at 6 and 15 months. Analysis of these data suggested validity problems for both uncorrected and corrected scores across the entire age range studied. The risks of overdiagnosis of retardation with uncorrected scores (false positives) and underdiagnosis with corrected values (false negatives) were underlined, and the need for new norms for premature populations was emphasized.

ALTERNATIVES TO TRADITIONAL NORMATIVE ASSESSMENT

In view of the many problems associated with the use of normative tests for the evaluation of high-risk infants, alternative formats for assessment are clearly needed. Current efforts in this regard are reflected in three strategies: reliance on clinical judgment, nontraditional use of normative instruments, and development of alternative techniques.

Clinical judgment has been most extensively employed in the primary health care setting. As the professional with the most extensive continuing contact with infants and their families, the pediatrician is uniquely suited to monitor the growth and development of young children. Traditional pediatric training programs, however, have not adequately prepared physicians for this role (Dworkin, et al., 1979). Most pediatricians do not employ standardized assessment techniques and rely instead on clinical judgment based upon a comprehensive history, physical examination, and office observation.

In a study of the customary clinical practices of 97 primary care pediatricians in five New England states, the use of standardized testing for developmental assessment was found to range from 19% to 38%, depending on the presenting concern (Shonkoff, Dworkin, Leviton, & Levine, 1979). Although the use of standardized instruments (primarily the DDST) was more frequently reported by younger physicians, the overwhelming majority relied on subjective, impressionistic assessment procedures. A number of systematic studies have challenged the contention that such impressions are an adequate substitute for standardized testing. Bierman, Connor, Vangel, and Honzik (1964) demonstrated that clinical judgment alone resulted in underdiagnosis of mental retardation in 2-year-old children, and Korsch, Cobb, and Asche (1961) observed that neither pediatricians' degrees of experience nor their stated confidence in making their appraisals were correlated with diagnostic accuracy.

An alternative to the use of clinical judgment alone that is employed by many clinicians involved in the assessment of high-risk infants is the nontraditional use of normative tests. In such cases, standardized instruments are employed in a nonstandardized manner. Some examiners use formal testing as a vehicle for systematic observations of young children. From these observations they generate descriptive formulations of competencies and deficiencies without calculating age levels or developmental quotients. Although such practices avoid many of the problems described earlier, the reliability and validity of the data obtained are highly variable and dependent on the clinical sophistication of individual examiners.

The deficiencies in current assessment practices argue strongly for the development and refinement of new evaluation techniques for infants and toddlers that are based on contemporary conceptualizations and clinical demands. The article by Dunst and Gallagher
The current dilemma of high-risk-infant evaluation can be summarized in the observation that researchers are attempting to use simple, traditional instruments to answer complex, contemporary questions. Traditional normative assessment tools were developed in a different era for different purposes and were based on a conceptualization of the developmental process that is very different from current thinking. When normative scales were first constructed, along with the concepts of developmental ages and quotients, arbitrary classification was needed to facilitate the segregation of abnormal children, and early intervention would have been characterized as an exercise in futility.

Current responses to atypical youngsters, on the other hand, demand qualitative, descriptive data that can be translated into individualized strategies for facilitating more optimal developmental progress. Numerical ratings, such as developmental quotients, generally provide little insight into current functioning, and are not necessary for planning specific intervention strategies, and are frequently equated with IQ scores by parents. Concepts such as "performance at the 17-month level" and "6 months of progress in the past 8 months" invite misinterpretation and may disproportionately affect the way parents perceive and consequently relate to their children.

When used to formulate a program of intervention, quantitative data based on normative testing often lead to stereotyped prescriptions oriented toward specific task mastery. Although there has been a virtual explosion in infant testing in the past decade, there has been inadequate critical analysis of the evaluation format and the utility of its products.

At a time when parental autonomy is eroding and professional intrusion into the infant-care giver relationship is becoming institutionalized, the interests of the developmentally vulnerable infant must be protected. Although the process of assessment has been the key to essential intervention services, it can also be the perpetrator of erroneous diagnostic impressions and inaccurate or premature labeling. Unlike its ancestors, a contemporary developmental assessment should be constructive and nonstigmatizing. Moreover, the right of environmentally handicapped or biologically disabled infants and their families to have specific disabilities identified and appropriately managed ought to be balanced by the right of nondisabled infants and their families to be left alone.

The combination of increased visibility of infants outside their families, growing public interest in early identification and intervention for children with developmental problems, and the limitations of available assessment techniques regarding long-term developmental predictions has created a sensitive dilemma: to develop and identify infants whose demands of the child are unlikely to be met.

The ability to abhor most infant abuses is evident. Research of learning, of abnormal in a species respect different their inf...
optimal ratings generally occurring, life interquated as "I" and "he" invite unilaterally and consequence of inter-normative ascriptive story. Al- plasm in there has the evaluation.

my is eroding the institutional mentality. Although the key to an also be diagnostic premature temporary construction, the right of biologi- family to and approved by the ear families visibility of owing public and inter-developmental of available long-term created a sensi- tive dilemma. The contemporary passion for better quantification and prediction in human development is formidable. The pressures to identify increasingly subtle disabilities of infants at an early age are growing, and the demand for more sophisticated measurement of the effects of early intervention programs is unlikely to diminish.

Although the pressures for greater accountability and precision should be applauded, it is most important that potential misuses and abuses of quantitative data in the measurement of human development be avoided. Researchers are now faced with the challenge of learning to distinguish between the concepts of abnormality and competitive disadvantage in a specific social system while continuing to respect the value of individual and group differences. They are beginning to temper their infatuation with scientific measurement by acknowledging that the process of human development is too complex to be easily reduced to strict numerical analysis.

The burden of many decisions in this area seems to be falling on the shoulders of the "experts." It is hoped that those who choose to accept these responsibilities understand the complexity of the issues and the need for careful study. The clinical use of normative evaluation techniques for high-risk infants, with all of their limitations and potential dangers, requires a sense of caution and humility on the part of the examiner. One must not be fooled by the apparent simplicity of administering and scoring a normative test. The concept of normality is ambiguous and never value-free. A critical look at how very young children are evaluated and labeled will tell researchers a great deal about society's evolving conception of humanity.

REFERENCES


Challice, S. Infants: Their chronological process. New Orleans Medical and Surgical Journal, 1887, 14, 893-902.


Rutter, M. *Infantische pages*.
Infants at Risk
PREDICTING THE DEVELOPMENTAL OUTCOME OF INFANTS AT RISK: GUIDELINES FOR INFANT ASSESSMENT

The evolution of our understanding of development as a transactional process influenced by biological and environmental factors has implications for the manner in which we measure infant competence. As Shonkoff reminds us in his overview of intelligence testing, the purposes for which we employ infant assessments are often quite different from the agendas for which they were developed.

The contributors to this issue on high-risk infants reiterate the limitations and misuse of infant intelligence tests to predict developmental outcomes. They describe recent research on infant behavior that leads to the new paradigms being used to develop ecologically valid assessments of competence that are more appropriate for evaluating atypical infant populations. Early psychometric tests enable users to label children as to whether they deviate from a linear path of development. These new assessments reflect current agendas for infant assessment, including (a) evaluation of atypical populations, (b) identification of subtle rather than gross problems, and (c) provision of data that can be used for programming early intervention.

Infants who are likely to attain skills at a slower rate than most children of the same age or who acquire skills that are less than efficient are considered to be "at risk." Trosen (1976) has identified three types of risk—established, environmental, and biological—categories that reflect environmental, as well as organismic contributions to development. Infants who are at established risk demonstrate early aberrant development, which is attributable to diagnosed medical disorders. Environmental risk results from deficient care-giving environments associated with suboptimal outcomes, and biological risk, although not based on a medical diagnosis, is associated with early biological insults related to subsequent aberrant development.

Measures of risk are reasonably accurate in predicting developmental outcomes for infants who lie at the extremes of these categories; however, we are much less successful in predicting subsequent outcomes for most of these infants. Our predictions are usually based on performance on standardized infant scales of mental and motor performance—measures most often standardized on populations from which atypical infants were excluded.

Because most of these assessments are global in nature, they fail to help us understand those early experiences that have highly specific effects upon the infant—information that is vital in order to program effective interventions. Wachs and Gruen's (1982) hypothesis that environmental parameters are specific to particular individuals, particular ages, and particular abilities directs us to our present task—to identify with greater precision the specific relationships that exist between organismic-environmental variables and the developmental outcomes of at-risk infants.

A major cause of our inability to predict the developmental outcome of children who experience biological or environmental insults is the nature of the research to date on early experience and development. The main effects and unidirectional models we have relied on are designed to identify single variables or dyadic interactions influencing subsequent behavior. Only recently have researchers begun to employ a transactional model that accommodates infants' contributions to their early development and environment and the diverse influences on their behavior.

Clarke-Stewart's (1977) longitudinal study of mother-child and father-child interactions is an excellent example of the kinds of studies needed to elucidate transactional patterns of interaction between characteristics of the child and the environment. As the infant development literature now informs us, infants respond differentially to their care-givers and influence adults' behaviors. Differential responses by adults are in turn discriminated by infants, who adjust their behaviors to adults' modifications.

Through this interactive process, infants clearly affect their own development. These interactions can be positive or negative. Frodi (1981), for example, found that the behaviors of abused infants differed from those of nonabused infants. The abused infants were more likely to be irritable, less sensitive to their mothers' cues, physically less attractive, or handicapped. Goldberg (1977, 1979) noted other infant characteristics that contribute to the infant's development, including gender, activity.
leveled, callfulness, responsivity, predictability, and demonstration of readable cues. When infant characteristics interact with parental and environmental variables, care givers may become abusive. Factors such as family history, marital harmony, economic status, education, and other family variables contribute to this complex picture.

Wachs and Green (1982) point out that the relationship between social-environmental variables and cognitive- intellectual development changes over time. As the main effects studies indicate, no single factor reliably predicts developmental outcome; however, when constellations of family and child variables are examined longitudinally, predictive accuracy increases. Examiners become more reliable, for example, by including information in their risk assessments culled from reciprocal exchanges.

The state of the art of infant assessment is changing in response to new research findings, current agendas for assessment data, and evolving paradigms. New detection techniques are being developed within the three risk categories delineated by Tosse, and predictive accuracy is increasing. The authors in this issue describe new directions in infant assessment. Regardless of the process or procedures they advocate, several guidelines are indicated.

The new information emanating from infant laboratories must be translated into formats that can be used in testing environments, and those environments must be congruent with the infant's life experiences. Infants can no longer be tested in isolation, and persons who influence and influence the infant must be included in research paradigms. Indeed, as Belsky and Tolan (1984) have indicated, we must examine more closely the effects of individual differences in family environments on infant development. They found, for example, that parental harmony was correlated with competent infant functioning.

In order to increase the predictive validity of assessments, we must use assessments that are time-bound, so that changes in the infant, family members, and family interactions can be monitored more closely to account for predictive changes. Assessment should also focus on the micro-level in order to identify specific targets for intervention. Measurement of competence on specific measures of infant motivation or on ordinal scales, for example, rather than on global psychomotor measures, will permit this fine-grained evaluation of the subsets of abilities that are compromised in general measures of intelligence.

REFERENCES


—Rebecca B. Fewell
Child Development and Mental Retardation Center
University of Washington
Seattle, Washington
Senior Editor