Cynthia Russo Silverman, Psy.D.
Psychologist, Boca Raton, Florida.

Grant Aram Killian, Ph.D.
Adjunct Professor of Psychology, Nova University, Fort Lauderdale, Florida.

William Joseph Burns, Ph.D.
Professor of Psychology, Nova University, Fort Lauderdale, Florida.

THE NEONATAL BEHAVIORAL ASSESSMENT SCALE

T. Berry Brazelton. London, England: MacKeith Press. U.S. Distributor—J.B. Lippincott Company.

Introduction

The Neonatal Behavioral Assessment Scale (NBAS; Brazelton, 1973, 1984) is an instrument that investigates the behavioral capacities of neonates at not less than 37 weeks gestation or more than 30 days after birth. The NBAS is designed as an interactive assessment of the normal, healthy, full-term newborn in which the examiner facilitates the best performance over a series of trials. In the process, the infant's inborn characteristics as well as his or her adjustment to labor, delivery, and new environment during the first month of life are assessed. The exam provides a means of assessing the newborn's attempt at homeostasis, the organization of inner and outer stimuli, as well as his or her sociability with caregivers. A major underlying consideration throughout the exam is the newborn's use of states of consciousness as a means of controlling impinging stimuli and regulating homeostasis. Specifically, the pattern of state changes is measured during the exam, and variability of state becomes a hallmark in a healthy neonate's adaptive functioning. According to Brazelton (1978), state becomes the infant's first line of defense against impinging stimuli. Obtaining best performance becomes a reflection of the baby's capacity to organize responses via state behavior.

T. Berry Brazelton, M.D., author of the NBAS, graduated from Columbia University School for Physicians and Surgeons in 1943 and accepted a medical internship in New York City. In 1945 he moved to Boston to serve as a medical resident at Massachusetts General Hospital before undertaking pediatric training at Children's Hospital. His interest in child development then led to training in child psychiatry at the Massachusetts General Hospital and James Jackson Putnam Children's Center from 1947 to 1950. Dr. Brazelton is presently Clinical Professor of Pediatrics at Harvard Medical School and founder of the Child Development Unit (a training and research center) at the Children's Hospital in Boston. He has published 18 books and more than 180 articles on child development. In 1989 Brazelton was appointed to the National Commission on Children by the presi-

dent, Senate, and House of Representatives, and he is a national cosponsor of Parent Action, a grassroots organization for parents.

The development of the NBAS came from the author's early work with infants some 30 years ago. Working as a pediatric intern, he noticed behavioral differences among neonates (Brazelton, 1978). The ability of the baby to shape the infant-caregiver dyad was tremendous. Based on the early works of Graham and associates in 1956 and Prechtl and Beintema in 1964 (cited in St. Clair, 1978), Brazelton utilized neonates' states of consciousness in the timing of maneuvers. The pattern of states appeared to be important characteristics of newborns during the period of adjustment to the extrauterine environment. The resulting formalized evaluation, originally titled the Cambridge Newborn Behavioral and Neurological Scales (Brazelton & Freedman, 1971), eventually became the Neonatal Behavioral Assessment Scale (Brazelton, 1973). The development of the scale involved the work of Freedman, Horowitz, Robey, Sameroff, and Tronick. The collaborators came from diverse backgrounds, including anthropology, pediatrics, developmental psychology, and psychiatry. Together with Brazelton, the group developed this comprehensive clinical interactive tool.

There is no comprehensive standardization study available to date for the NBAS. Silverman (1991) has argued for additional reliability studies as well as long-term predictive validity data. The importance of such studies cannot be ignored. According to Kline (1986), test scores from clearly defined samples representative of the population at large are the only basis for comparison. Consequently, if individuals are to compare test scores, only when there is a standardized sample large enough can we trust the results from various research projects. The variability from site to site might be due to cultural differences rather than differences in babies. Because the newborn is influenced by various interuterine and environmental factors, Als, Tronick, Lester, and Brazelton (1977) question the meaningfulness of collecting normative data on such a sample to be used for comparative purposes. They believe it is more important to examine the behavior of the newborn from a particular culture rather than to imagine a baby's behavior might be representative of the U.S. population on the whole. Needless to say, both researchers and clinicians who use the NBAS are forced to adapt their method to a scale not undergoing standardization procedures.

In an effort to explore the effect of background variables of neonatal behavior on a sample of 221 normal newborns, Lancioni, Horowitz, and Sullivan (1980) completed a study utilizing the Neonatal Behavioral Assessment Scale with Kansas Supplements (NBAS-K) and believe the NBAS-K provided for a more individualized infant assessment based on modal or average scores representative of the typical performance across newborns rather than the best performance measured with the NBAS. Their analyses of data are from the first 3 days of life with a subset of 106 infants later retested at 2 weeks and again at 1 month of age. The background variables included age and education of parents, number of previous births, number of pregnancies, length of labor, time since last feeding, birth weight, and current weight at time of exam. The results of those analyses suggest that the background variables selected had minimal influence on performance. Interpretations of the test results are inadequate, however, if one uses the criteria suggested by Kline (1986) recommending a sample size of 500 to reduce statistical error and stratified sampling.

The NBAS has undergone five revisions in the last 20 years. A preliminary version was developed in 1971 by Brazelton and Freedman (the previously mentioned Cambridge Newborn Behavioral and Neurological Scales) but was never published. The administration of items and a definition of neonatal behavioral states and elicited responses were described in an unpublished manual, which gave suggestions for grouping items in a meaningful fashion for purposes of interpretation.

In 1973 Brazelton and his colleagues published a more comprehensive version of the scales through *Clinics in Developmental Medicine.* Entitled the Neonatal Behavioral Assessment Scale (NBAS), this instrument overlapped the 1971 version with a few important exceptions. The 1973 revision suggested an order of administration of the items and general procedure, taking into account the appropriate states in which the assessment of each behavior can be made. Photographs clearly depicting the neonate in six states were included: a) deep sleep, b) light sleep, c) drowsy, d) alert, e) eyes open, and f) crying. Scoring definitions, examiner training requirements, and early research with the scale also were described. The 1973 edition consisted of 27 behavioral items and 20 reflex items intended to score a full-term (40 weeks gestation) normal Caucasian infant as he or she relates to inner and outer stimuli. The 27 NBAS behavioral items, which measure the infant's interactional capacities with his or her caregiver as scored on a 9-point scale, include response decrement to light, rattle, bell, and pinprick; orientation to inanimate visual, inanimate auditory, animate visual, animate auditory, and combined animate visual and auditory stimuli; alertness; general tonus; motor maturity; pull to sit; cuddliness; defensive movement; consolability; peak of excitement; rapidity of buildup to a higher state of arousal; irritability; activity; tremulousness; startle; lability of skin color; lability of states; self-quieting activity; hand-mouth facility; and number of smiles.

Sullivan modified the NBAS and called it the Neonatal Behavioral Assessment Scale with Kansas Supplements (NBAS-K). He asked trained examiners to score infants' behavior in a standard manner looking for best performance and to score their "modal" or average scores on consolability, orientation, and defensive maneuvers (Horowitz, Sullivan, & Linn, 1978). In addition, Sullivan wrote scoring definitions for five additional scales: orientation to inanimate visual and auditory stimuli, quality of infant's alert responsivity, examiner persistence, general irritability, and reinforcement value of infant's behavior. Horowitz et al. (1978) reported that the NBAS-K provides for a more individualized infant assessment based on the modal or average scores achieved with healthy full-term babies. For a modal score to be assigned, at least three to four trials of an item is preferred (Brazelton, 1984).

Investigators conducted research on the stability and structure of the NBAS-K to determine if modal scores were more predictive of a newborn's performance than best scores. A study conducted by Lancioni et al. (1980) reported modal scores did not demonstrate higher correlations over the first 3 days of life than best scores. In addition, there are times when a modal score is difficult to obtain. Disorganized neonates may not be accessible for repeated item administrations. Best scores reflect the neonate's attempt at organization of inner and outer stimuli as well as the amount of facilitation necessary to produce optimal behavior (Brazel-

ton, 1984). Modal scores may be useful as additional information in determining the newborn's performance.

In 1982, Als, Lester, Tronick, and Brazelton developed an instrument to systematically assess the behavior of the preterm infant and document patterns of developing organization. The development of the scale was based on the belief that preterm infants developed quite differently from full-term infants. The new research instrument, called the Assessment of Preterm Infants Behavior (APIB), assessed the neonate when he or she was without medical aids, in a room temperature of 72–80 degrees Fahrenheit. The rationale for developing the APIB was that preterm infants had difficulty maintaining the state of alertness necessary for appropriate administration of the NBAS. In addition, the criteria for scoring term infants' responses were never appropriate for preterm infants. The latter demonstrate an inadequate intake of stimuli, a motor system with unsuccessful inhibition of responses, and autonomic nervous system exhaustion. Detailing changes over a relatively short period of time is helpful in understanding the preterm infant's communication and may lead to adaptations in the environment. Interested readers should refer to Als et al. (1982) for more detailed description of the instrument.

Despite the fact that the ABIP is more time-consuming than the NBAS, it has the advantage of being appropriate for preterm or term infants. Because the NBAS items are included within the ABIP, the two scoring systems can be compared in the same infant. Burns, Deddish, and Hatcher (1982) studied the usefulness of the NBAS with preterm infants at 33 to 40 weeks postconception. Their results support that the NBAS is not an appropriate scale for use with preterm infants. Although the scale did reflect the babies' immature organizational systems, there was a need to adjust the administration of items as the result of decreased responsiveness. These authors thus argue that the NBAS should not be used as a clinical instrument for preterm infants. The APIB is designed to assess the way infants attempt to differentiate and modulate their systems, and it offers a complete description of the specific behaviors these infants use in their efforts to maintain stability. "The continuous identification of the thresholds of balance and stress is the key feature of this examination" (Als, 1984, p. 31).

According to the manual, the second edition of the NBAS (Brazelton, 1984) was developed specifically to a) restate the original purposes of the scale, b) clarify the definitions, c) embellish the training techniques, d) summarize the current research applications, and e) add an inanimate visual and auditory item to the basic scale. Nine supplementary items are included in the 1984 addition and deemed an optional tool, especially with fragile babies. These items, based on APIB items and the NBAS-K, include a) quality of alert responsiveness, b) cost of attention, c) examiner persistence, d) general irritability, e) robustness and endurance, f) regulatory capacity, g) state regulation, h) balance or motor tone, and i) reinforcement value of the infant's behavior. An additional set of modal scores have been added to the 1984 version on items such as orientation, consolability, and defensive maneuvers. Horowitz et al. (1978) argue that modal or average scores can predict a baby's functioning under a normal home situation better than can "best performance" scores. In both editions of the NBAS manual, however, Brazelton and his colleagues recommend using "best performance" scores as the most accurate

reflection of infant status. Because a newborn infant is in the midst of an adjustment to birth, the use of optimal rather than modal performance scoring is preferred. "Best performance" scores are interpreted in the context of the amount of effort the examiner and newborn invest in their interactions with each other during the administrations. Although the scores were not designed to predict functioning under home conditions, these scores are considered a good estimate of developmental newborn status. Therefore, modal scores may be used as additional information alongside optimal scores as a way of assessing functioning at home.

The 1984 NBAS package contains an audiovisual tape by Brazelton ("Neonatal Behavioral Assessment Scale"), a test manual that includes the basic background for the NBAS, scoring guidelines, and a data summary sheet, and a set of guidelines for using the NBAS with newborns and their families. There is also a testing kit of equipment necessary to complete the assessment: a) a shiny red ball, b) a flashlight (batteries not included), c) a rattle, d) a bell, e) a tactile probe, and f) a portable carrying case. The package also contains eligibility requirements for professionals wishing to become certified as NBAS-trained examiners and a list of NBAS training centers. The manual provides a description of the test, its development and uses, reliability and validity information, normative data, instructions for administration, scoring, and interpretation, methods for data analysis and prediction studies, suggestions for using the NBAS in research, criteria for selecting scale scores, descriptive criteria for identifying states, and photographs of typical infant responses to NBAS items.

The 20 NBAS reflex items are each scored on a 4-point scale from 0 to 3. Most healthy full-term babies obtain an average score of 2 on these items. The tonic neck reflex, ankle clonus, and nystagmus may receive either a 0, 1, or 2, all considered within normal limits. Brazelton suggests if three elicited scores are deviant, the infant should be referred for a complete neurological exam following the NBAS. There are two global behavioral scales—attractiveness and need for stimulation. The attractiveness dimension measures the infant's organized response capacity, integration of behavior, and positive feedback to the examiner. These factors contribute to the neonate's social attractiveness to the extrauterine environment. "Need for stimulation" refers to the baby's use of and need for stimulation to help organize his or her responses. Both items are rated 0 to 3, with 0 being infants in need of external help to organize and 3 measuring optimal self-organization. The NBAS now contains 28 behavioral items instead of the earlier 27. All of the original behavioral and reflex items are included in the 1984 version.

The manual suggests that scoring can be done at the end of the administration to minimize the examiner's distractibility from the overall flow of infant responsivity. The exam itself usually takes about 20 to 30 minutes. There is a suggested order of administration, although Brazelton suggests that one must follow the infant's availability in order to assure that he or she gives the best performance. Throughout the manual and on the scoring sheets, the state required for each item to be administered is listed in parentheses. Examiners must follow the proper state for each item administered or delay the item if the infant is not in the necessary state. After the administration is completed, scoring usually adds about 15–20 minutes to the total exam. The examiner makes a check mark in the appropriate box on the score sheet for the 28 behavioral items (using a 9-point scale).

There are some difficulties with the interpretation of scoring, and the scale does not have a uniform scoring system. Some items are considered optimal at 9, while others are optimal scored at 5. No summary score can be derived because items are deemed optimal at different points. Instead, clustering for data analysis those items that interact in similar ways and describe global functions in the baby takes place (Lester, Als, & Brazelton, 1982). The clusters are a) habituation, b) orientation, c) motor performance, d) range of state, e) regulation of state, f) autonomic regulation, and g) reflexes. It is also suggested that the examiner write a paragraph at the end of the behavior scoring sheet overall performance, major state fluctuations, and any additional comments relative to the assessment.

The NBAS can be used in a number of settings (hospital, outpatient clinic, home setting), although environmental conditions such as noise level and room temperature must remain standardized. Any deviations from these conditions should be reported on the score sheet. The exam begins ideally when the infant is asleep, midway between feedings. At least two exams on different days are recommended in order to avoid making conclusions on the basis of chance variations. Examiners are more likely to draw erroneous conclusions on the basis of a single exam. Administration takes place ideally on day 3 or 4, when the infant is free of the stress of delivery, and again on day 9 or 10, when the baby is at home. Brazelton (1978) points to "the maximum predictive validity of the scale in the pattern of recovery reflected in repeated assessment over the first few weeks of life rather than in any one assessment" (p. 8). He goes on to say that the scale was not designed for neonates of less than 37 weeks gestation or more than 30 days old. At 1 month after birth, some of the reflexes have all but disappeared.

Detailed information regarding precise methods for administration are clearly outlined in the manual and presented in a user-friendly fashion. The examiner plays an integral part in the administration of the NBAS, as he or she is the extrauterine environment to which the neonate responds. Because the score is based on clinical judgment, examiners must be trained in order to make these judgments. Brazelton has recommended that all researchers and clinicians, regardless of professional background, participate in the training at one of the three specified training centers before using the scale in either a clinical or research project. After an initial acquaintance with the use of the scale (assessing and scoring), new examiners are advised to participate in a 2-day training session to ensure interrater reliability. It is quite time-consuming to learn to administer the NBAS and rather expensive (at the time of this writing, the fee for the Brazelton-Based Consultation with Infants and Families Training Program is $950.00; if the trainer travels to the examiner's site, travel and lodging expenses are additional costs to the trainee).

Practical Applications/Uses

According to the manual, the Brazelton Neonatal Behavioral Assessment Scale is used primarily in research settings and in clinical practice as an interactional tool between newborn and examiner. However, despite the author's warning that the NBAS is not a normative test for infant development, the scale continues to be used as a test. For example, it is used in 500 research centers in the United States, 15

in Europe, and 10 in Asia to help parents understand and attach to their new babies. "The NBAS is not, at present, a tested clinical tool" (Brazelton, 1984, p. 103). Nevertheless, the NBAS does offer a trained examiner the descriptive status of a neonate. Full-term newborns respond in such a way during the exam that their observable physiological reactions, social responsivity, and motor and cognitive skills are scored as reflecting their ability to meet the demands of the environment.

Several studies involving cross-cultural comparisons with the NBAS have been reported, and the search for developmental differences at birth within different cultures has been cited (Brazelton, Koslowski, & Tronick, 1971; Keefer, Tronick, Dixon, & Brazelton, 1982; Kestermann, 1981). Numerous studies have focused on prenatal and perinatal factors associated with NBAS scores in different parts of the world (Brazelton, Tronick, Lechtig, Lasky, & Klein, 1977; Dunlop, 1982; Saco-Pollitt, 1981), exploring factors such as maternal nutrition and parity, drug condition during labor and delivery, and correlations between birth weight and attitude. Certain cross-cultural research has investigated nutrition and teenage childbearing in Puerto Rico and the United States (Lester, Garcia-Coll, & Sepkoski, 1982), and some has compared a typical pattern of fetal growth among normal-size infants born to teenage and older mothers in Puerto Rico (Lester, Garcia-Coll, Valcaicel, Hoffman, & Brazelton, 1986). Included in the latter study was the use of supplementary NBAS items. The findings suggest that the NBAS supplementary items do provide qualitative aspects of neonatal behavior based on the examiner's subjective impressions, but that they do not add predictive information for groups of at-risk infants. Overall, it is important to note that the NBAS itself may not be sensitive to the performance of at-risk infants because it was designed for assessing full-term babies (Brazelton, 1973, 1984).

The NBAS studies of Gomes-Pedro, Bento de Almedia, Silveria da Costa, and Barbosa (1984) and Choi and Hamilton (1986) have investigated, in Portugal and Korea, respectively, the effects of early contact on the behavior of newborn infants and on maternal attitudes towards the newborn. These studies illustrate that early contact with newborns facilitates caring maternal feelings in mothers toward them.

Although sample sizes in all of the studies discussed have been criticized as less than adequate (Kline, 1986) and limited to particular areas in a given country, many sources of variability in neonatal performance have been explored. There is a need to utilize a broader sample of neonates from different cultures in order for valid conclusions to be drawn. Based on the notion that the standardization of context, such as room temperature and noise level, has been a critical concept in Brazelton's (1973, 1984) instruction manual, DeVries and Super (1978) have selected those items on the NBAS that are most sensitive to contextual influences and argue that caution should be used when drawing conclusions about cross-cultural research. In particular they believe that there are differences in the standardization of testing procedures based on differences in contextual variations. They studied the effects of home versus hospital, and tested infants and maternal beliefs about neonatal vulnerability within four regions of Africa. They propose that the NBAS findings obtained in cross-cultural research may be better viewed as indications of cultural bias rather than infant performance. Brazelton does not address this issue in the 1984 NBAS manual.

The effects of prenatal alcohol exposure and maternal narcotic addiction has been documented widely. The NBAS has been used to assess newborn performance in a number of studies. Early investigations of the effects of neonatal narcotic withdrawal suggest that the Brazelton scales are sensitive to many signs of narcotic withdrawal. Addicted newborns showed deficits in state control and interactive ability with an increased irritability and tremulousness in response to a stimulus (Chasnoff, Burns, & Schnoll, 1983; Kron, Kaplan, Finnegan, Litt, & Phoenix, 1975; Soule, Standley, Copans, & Davis, 1974; Strauss, Lessen-Firestone, Starr, & Ostera, 1975; Strauss, Starr, Ostera, Chavez, & Stryker, 1976). Babies who were delivered by mothers using PCP have been found to show significant changes on the NBAS compared with mothers withdrawing from polydrug use (Chasnoff, Burns, Hatcher, & Burns, 1983). Classic signs of withdrawal were not found in polydrug exposed babies, although significant changes in levels of consciousness, hypotonicity, and hyperreflexia were reported (Chasnoff, Burns, Hatcher, & Burns, 1983; Chasnoff, Hatcher, & Burns, 1982; Chasnoff, Schnoll, Burns, & Burns, 1984; Rogan et al., 1986). Methadone-exposed infants measured with the NBAS exhibited deficits in interactive behavior within the first 3 days of life (Finnegan, 1984). In a study by Jeremy and Hans (1985) involving methadone-exposed neonates, NBAS scores were reported during the first week of life and again at 4 weeks. Findings during the first week of life are consistent with other reports in the literature; by 1 month, the problems in neurobehavioral functioning largely have disappeared. The study suggests that methadone does not continue to interfere with neurobehavioral functioning on a long-term basis.

The NBAS also has been utilized in the evaluation of the effects of cocaine on pregnancy and the newborn (Chasnoff, Burns, & Burns, 1987; Chasnoff, Burns, Schnoll, & Burns, 1985; Chasnoff, Griffith, MacGregor, Dirkes, & Burns, 1989). On the NBAS, infants exposed to cocaine exhibited significantly more irritability, tremulousness, and state control as compared with a control group or with infants whose mothers used methadone. One study (Chasnoff et al., 1987) reported a significant increase in the rate of sudden infant death syndrome for cocaine-exposed infants as compared to narcotic-addicted infants.

Additionally, the NBAS has been used in studies relating intrauterine alcohol exposure to infant behavior with special emphasis on the response decrement items (Ernhart et al., 1985; Streissguth, Barr, & Martin, 1983). Varying amounts of alcohol ingestion during different periods in gestation were studied as well (Coles, Smith, & Falek, 1987; Coles, Smith, Fernhoff, & Falek, 1985; Coles, Smith, Lancaster, & Falek, 1987; Smith, Coles, Lancaster, Fernhoff, & Falek, 1986).

Other research has used the NBAS to compare the behavioral responses of infants whose mothers smoked 15 or more cigarettes a day compared to those of nonsmoking mothers (Saxton, 1978). The study failed to show statistically significant difference in overall infant behavior. One limitation in using the NBAS is that there are frequently numerous interpretations of a specific score and scoring is somewhat suggestive of these meanings. Studies examining caffeine ingestion during pregnancy have shown that lower levels of caffeine may affect the infant when socially interactive demands are made. Increased caffeine levels may produce spontaneous and noninteractive behaviors (Emory, Konopka, Hronsky, Tuggey, & Dave, 1988).

Research with neonates exposed in utero has significant implications for future mother-infant interactions. Tronick (1987) discusses some limitations of the Brazelton scale in respect to such research, offering the opinion that NBAS scores are not reflective of any one specific behavior but rather an integration of multiple causative factors. He goes on to suggest that additional assessments be utilized in the study of teratogenic effects on the newborn as well as repeated NBAS exams.

Numerous studies have been undertaken to document the scale's usefulness as an interventional tool for educating parents and health care personnel about newborn behavioral abilities. Several suggest that teaching parents about the abilities of their newborns through a demonstration of the NBAS fosters early positive interactions (Anderson & Sawin, 1983; Beal, 1989; Golas & Parks, 1986; Meyers, 1982; Szajnberg, Ward, Krauss, & Kessler, 1986; Widmayer & Field, 1981; Worobey & Belsky, 1982). However, Belsky (1985) cautions that some of these studies have used confounding interventions such as videotaping, which may have created results due to experimenter effects (Widmayer & Field, 1980). He goes on to suggest that extra contact with the examiner also may have contributed to the positive findings. Beal (1986) cautions restraint as well when reviewing research that implies long-term beneficial effects of early intervention with the NBAS, and she reviews those studies that demonstrate short-term positive effects of NBAS demonstrations during the first year of life. Worobey and Brazelton (1986) suggest that one NBAS-based intervention may not produce a lasting positive effect on a family, but believe it offers an important first step toward developing a feeling of mastery in the parents. It is the opinion of Nugent and Brazelton (1989) that interventions with the NBAS during the early period of newborn development will enhance a positive reciprocal communication in the family with long-term consequences. They suggest that psychological benefits derive from utilizing the NBAS within the family system.

Infant temperament also has been the subject of NBAS research. Such characteristics as infant irritability, activity level, sucking, eye contact, and so on have been thought to influence caregiver behavior during the newborn period and have been assessed using the NBAS (Aleksandrowiez & Aleksandrowiez, 1974; Crockenberg, 1981; Crockenberg & Acredolo, 1983; Crockenberg & McClusky, 1986; Jones & Lenz, 1985; Osofsky, 1976; Osofsky & Danzger, 1974; Penman, Meares, Baker, & Milgrom-Friedman, 1983; Worobey, 1986a, 1986b). These investigators found temperament to be related to NBAS findings. However, the generalizability of many of these studies are limited by small sample size, inclusion of high-risk infants, and possibly a general patterning of maternal perception rather than a true measurement of inherited traits.

The role of obstetric medication on neonatal behavior has been evaluated by the NBAS in a number of studies (e.g., Aleksandrowiez & Aleksandrowiez, 1974; Kuhnert, Linn, Kennard, & Kuhnert, 1985; Muhlen, Pryke, & Wade, 1986; Standley, Soule, Copans, & Duchouny, 1974). Although the studies report neonatal effects for the heavily medicated group of infants as measured by the NBAS, there are methodological difficulties. A wide range of drugs is reported within each group, as well as varying susceptibility to drugs among the mothers tested. Subject loss also posed a problem with this research: Babies could not be assessed because they were taken to special care nurseries. The authors agree that behav-

iors produced by obstetrical medication may affect the interaction between infant and caregiver, which may in turn affect subsequent infant development.

Several investigators have attempted to study the effects of tactile stimulation on the behavior of preterm infants. The specific effects of stimulation have varied across studies and must be viewed with caution. As noted, the scale was designed to be used with full-term infants (Brazelton, 1973, 1984), and Brazelton believes that the supplementary items should be used when assessing preterm infants with the NBAS. Burns et al. (1982) conducted a study using the NBAS to determine the response characteristics of preterm infants born at 32 to 40 weeks postconceptual age. Their results indicate that the NBAS without modification presents an inappropriate tool with preterm infants. Perhaps the APIB is the better scale for such studies, as it was designed for preterm infants.

Despite the good evidence to the contrary, many researchers have chosen to use the NBAS with preterm infants. Because of the small sample size in these studies, it is very difficult to interpret the results (see Field, 1979; Leijon, 1982; Paludetto et al., 1982; Paludetto, Rinaldi, Mansi, Andolfi, & Del Giudice, 1984; Scafidi et al., 1986). Scanlon, Scanlon, and Tronick (1984) attempted to describe the behavior of the extremely premature neonate using a modified version of the NBAS. These investigators state that the examination was a forerunner of the APIB, and the use of the APIB is supported with preterm infants. Forty-five infants were enrolled in the study. Als (1986) describes the theoretical framework underlying the development of the APIB and its appropriateness for use with the preterm infant as well as at-risk infants and healthy full-term infants.

Regarding special populations of high-risk infants, NBAS studies reported include hydrocephalic infants, jaundiced infants, the effects of maternal anemia on infant behavior, pulmonary hypertension in the neonate, preterm infants with intracranial hemorrhage, neonates with seizures, infants of multi-risk families, and full-term but underweight infants (Als, Tronick, Adamson, & Brazelton, 1976; Anderson et al., 1989; Aylward, Lazzara, & Meyer, 1978; Emory, Tynam, & Dave, 1989; Escher-Graub & Fricker, 1986; Frances, Self, & McCaffree, 1984; Greene, Fox, & Lewis, 1983; Hofheimer, Poisson, Strauss, Eyler, & Greenspan, 1983; Murphy, Scher, Klesh, & Guthrie, 1988; Osofsky, 1974; Vaughn, Brown, & Carter, 1986).

The NBAS has been used in nursing practice to identify infants that have shown suspect behavior during the administration of routine nursing care similar to the NBAS assessment. The NBAS results have been consistent with nursing observation of infant behavior (Maloni, Stegman, Taylor, & Brownell, 1986). The NBAS also has been used as a teaching technique for neonatal nurses, who then can educate parents regarding their newborn's capacities (Gibbs, 1981).

Several studies have documented maternal reactions to various modes of delivery and the impact of mother-infant interactions following delivery. No significant differences were found for mother-infant interaction regardless of delivery method (Gottlieb & Barrett, 1986; Kochanevich-Wallace, McClusky-Faucett, & Meek, 1987; Leijon, Finnstrom, Hedenskog, Ryden, & Tylleskar, 1979). Caution should be exercised when interpreting these results, as sample sizes are less than adequate in all of the studies.

Mental health personnel and researchers would be the most likely candidates to administer the NBAS. Administration and scoring of the scale is open to any

individual who has undertaken the training program set up by Brazelton and his colleagues. Scoring can take up to 15 minutes at the end of the exam. As previously described, the examiner is an interactive participant in the evaluation, attempting to draw out the neonate's best performance. Because the scale does not set a fixed order of administration, the examiner must be sensitive to the infant's cues. Brazelton (1984) reports that examiners can be trained to an 85%–90% level of interrater reliability agreement, which can be maintained for up to 2 years without retraining. The scoring criteria presented in the manual are explicit. Brazelton (1984) recommends that the trainee work with at least 20 to 25 babies before training begins in order to master the ability to understand neonate behavioral cues.

Clopton and Martin (1985) argue that the reported percent of interrater agreement may be inflated due to four problems: 1) the predominant use of a few scores, such as midway scores, and the rare use of others, such as extreme scores of 1 or 9; 2) the practice of allowing a difference of 1 point to be counted as agreement, when several items function as 3-point scales; 3) the subjective administration of the decrement items; and 4) the fact that translation to a prior scale does not produce high interrater reliability. Clopton and Martin believe interrater reliability should be reported as exact agreements and that the actual percentage of agreement between trained observers is probably less than 85%. Interrater reliability is an important feature of a good test, and Brazelton does not address the issue of this type of inflation in the manual.

The manual suggests three training films to view as part of the training phase. These films describe the scale in detail and illustrate the performance of various infants. The manual also recommends that the trainee witness a live demonstration by a certified examiner. There is a 2-day reliability training workshop as well, which is conducted by a certified trainer from one of the seven training centers. This phase of training, which serves to bring the trainee to the percent level of reliability agreement, presents a discussion of administration and scoring issues, a demonstration by the trainer, and joint scoring by both examiner and trainee. The trainee must administer the scale to two babies while the examiner and trainee independently score the scaled items. The manual cautions that no more than two babies should be examined in one day since observational powers tend to diminish as fatigue sets in. The examiner must test the infant at least twice over the first few weeks of life in order for a pattern of recovery to be established, which is more meaningful than a single examination.

The 1984 NBAS manual describes the order of administration in great detail. Although there is no fixed item sequence, as noted previously, grouping the items as suggested ensures a sequence of stimulation for the baby that in turn reflects his or her coping mechanism in the face of increasing stress.

The instructions for scoring each item are presented clearly in the manual. Scoring is done by hand immediately following the exam. (There is, however, a computer-generated scoring method for the Lester, Als, & Brazelton [1982] cluster system.) With 48 items, scoring can be quite cumbersome. The individual item scores are not independent, nor do they have equal intervals. Further, NBAS items cannot be transformed into traditional summary scores, although a number of statistical analyses have been performed over the years. Popular approaches have

been item comparisons, reduction of the items to several subscales, factor analytic approaches, a priori clusters, and Lester et al.'s (1982) cluster system. For an excellent review of the various scoring systems, see Als et al. (1982). They suggest that there are three drawbacks to the item-by-item comparisons: the items are not independent, the differences arising in a post hoc analysis may be due to chance, and the reported differences may not be greater than the reliability of the examiner. In looking at the reduction of items to various subscales, the main disadvantage is that the items cannot be combined because of differing optimal ranges of performance. In factor analytic methods, sample sizes are often too small and factors are compared at different points in time, a problem because the factors may not be measuring the same variables. Als et al. (1982) recommend a priori clustering of scale items in four areas (physiological organization, motoric processes, organization processes, and interaction organization) and scaling functioning along three points (superior, average, and worrisome). Judgments in scoring here are based on clinical experience with the scale, "not . . . on a frequency distribution of a 'representative' sample" (Als et al., 1982, p. 19).

A priori clustering does provide a description and assessment of overall behavioral competencies in the newborn and adequately deals with the problems of nonlinearity of items (Als et al., 1977). However, Maier et al. (1983) argue that a priori scores are not sensitive to subtle differences among neonates as the result of a limited 3-point scale. Brazelton (1984) feels strongly that babies should not be given a single summary score because they are in a state of rapid change. He recommends the use of Lester et al.'s (1982) scoring method, which clusters the individual NBAS items to describe seven global functions. Their method combines the 28 behavioral items into six clusters and the seventh is the total number of deviant reflex scores. The seven clusters are habituation, orientation, motor performance, range of state, regulation of state, autonomic regulation, and reflexes. Curvilinear scale items are recorded as linear. Gyurke (1988) conducted research that compared the scoring procedures of the Als (1978), Lester et al. (1982), and Jacobson, Jacobson, Fein, and Schwartz (1984) clusters. "Both Als' and Lester's systems recode midpoint optimal scores to high scores without reference to empirical data" (Gyurke, 1988, p. 203). In comparing the three methods, Gyurke found that all three scoring approaches detected differences on items assessing motor maturity and orientation clusters.

Technical Aspects

The disadvantage of the Neonatal Behavioral Assessment Scale continues to be its lack of adequate representative normative samples and standardization scores with statistical properties. There have been no comprehensive standardization studies published to date. Interpretation is based on clinical judgment concerning the behavior that makes up a given score; thus clinical experience with babies and a knowledge of infant development are necessary tools for understanding the meaning of a given score.

Although Brazelton never intended the NBAS to be used as a test, it is defined as one according to the three categories of test instruments outlined in the *Standards for Educational and Psychological Testing* (American Educational Research

Association, American Psychological Association, & National Council on Measurement in Education, 1985): "constructed performance tasks, questionnaires, and to a lesser extent, structured behavior samples" (p. 4). The NBAS falls into the category of structured behavior samples and therefore must be considered a test.

Standardization in the information given to test users is a necessity if reasonable decisions are to be made about test performance. Because the NBAS is defined as a test according to the *Standards*, adherence to the rules that define construction and evaluation of tests is a major consideration. The NBAS fails to provide the standardized information needed for reasonable comparability of infants' performance. Despite the fact that it was never designed as a test, this review will consider its psychometric properties because it falls under the AERA/APA/NCME definition of a test.

According to Standard 6.3, "when a test is to be used for a purpose for which it has not been previously validated, or for which there is no reported claim for validity, the user is responsible for providing evidence of validity" (AERA, APA, & NCME, 1985, p. 42). Brazelton and his colleagues have clearly stated in 1973 and 1984 manuals that the NBAS is not intended to be used as a test. Test users, therefore, must bear the responsibility for using the NBAS inappropriately.

A psychological test must be reliable and valid if it is to be described as a good test (Kline, 1986). Reliability is a necessary but not sufficient prerequisite for validity. The NBAS has questionable reliability and validity based on the AERA/APA/NCME standards. The two methods reported in the manual and utilized in the studies under review address test-retest reliability and interrater reliability. The characteristics of a reliable test are that it yields the same score for individuals over time (test-retest reliability) and that two independent examiners observe the same stimuli and scores consistently (interrater reliability or agreement). High test-retest reliability is associated with good predictive validity because the observed behavior is viewed as stable across time (Jacobson et al., 1984). The test-retest reliability for the NBAS utilizing the Pearson *r* with repeated exams report low to moderate day-to-day stability (Horowitz et al., 1978; Kaye, 1978; Sameroff, 1978). The Pearson *r* reflects the relative standing of an individual score within a group on two successive occasions. In a sample of 44 Kansas and Israeli infants, Horowitz et al. (1978) reported correlations from .20 to .50. Sameroff (1978) and his colleagues demonstrated that individual items, factors, and clusters of newborns tested over two sessions demonstrated instability across 2 days. Their analysis was based on a sample of 35 infants who were tested on two consecutive occasions. Kaye (1978) tested 42 neonates at Day 2 and again at 2 weeks of life to explore the stability of various factor scores by "smoothing" them over several sessions. He averaged normalized scores over related items and then averaged over several independent exams. In addition, he obtained scores from the Boston Lying-In Hospital. Kaye concluded that there were no significant correlations between any factors across administrations.

In the 1984 manual, Lester (1984) discusses the method of assessing test-retest reliability on the NBAS. It serves to identify the number of items measuring stable or variable performance in an individual across exams rather than assessing an individual's score within a group on repeated exams. Horowitz and Brazelton (1973) reported data on 60 neonates tested on the third to fourth day of life and

again at 1 month of age. Test-retest reliability was computed by using the number of items for which there was an agreement divided by the number of agreements plus the number of scores for which there was a disagreement. Two scores that were identical or within 1 scale score point of each other were considered an agreement. A second level was calculated by determining if two scores were within 2 points of each other, in order to consider low reliability as a function of wide disagreements. A mean test-retest stability was reported at .592 for the first level and .783 at the second level. Nunnally (1978) believes reliabilities of .70 or higher will suffice if basic research is being conducted. Brazelton's test-retest reliability measures do not meet these standards at the first level according to Nunnally.

In the 1984 manual, Lester (1984) suggests that the standard psychometric properties of the Pearson *r* will yield low to moderate test-retest reliability because the method is not a useful measure of day-to-day stabilities in an individual neonate's performance across time. Brazelton (1984) acknowledges that the test-retest characteristics of the NBAS are poor, but he does not offer another measurement of reliability in the manual that would be sensitive to sources of error, including subjectivity of scoring and personality maturation over time. Although many studies under review have been conducted to estimate the test-retest reliability of the NBAS, Nunnally (1978) states that test-retest designs can be used only if three criteria are met: the repeated measurements are independent, no significant development occurs between the tests, and the temperament of the neonate does not change over time. Considering the tremendous growth and development within the neonatal period, another means of assessing reliability should be considered. Nunnally suggests using measures of internal consistency that are based on the average correlation among the sampling of situational factors that accompany the administration of items. He believes the items should correlate highly with one another. Internal consistency measured by coefficient alpha is necessary to support the ideal that clusters include items strongly correlating with each other. As Nunnally (1978) points out, internal consistency should be the major aim of all test constructors.

Jacobson et al. (1984) reviewed the three methods most frequently used to reduce NBAS data: factor analysis (Sameroff, 1978); clusters (Als, Tronick, Lester, & Brazelton, 1979); and Lester clusters (Lester, Als, & Brazelton, 1982). Their study analyzes the problems related to each method. Because in the 1984 manual Brazelton and his colleagues strongly recommend using the Lester clusters, it is important to review the data here. Jacobson et al. found while reviewing the Lester clusters that every item is included in each cluster, including items not found to relate consistently to the others (e.g., consolability). In addition, as the authors point out, the optimal midpoints were chosen without reference to objective empirical criteria. Although the Lester et al. approach offers advantages over the other methods, the psychometric properties of these clusters have not been reviewed extensively with respect to reliability or validity. Jacobson et al. evaluated the Lester clusters, comparing their within- and between-cluster correlations in a sample of 162 newborns. Motor items and range of state showed little internal consistency when recorded as Lester et al. recommend. Jacobson et al. constructed a revised set of clusters from NBAS data that exhibited greater internal consistency

(see the Jacobson et al. article for an excellent review of the distributional and psychometric properties of the clusters). These latest findings should be included in a revised edition of the manual.

Osofsky and O'Connell (1977) conducted a study based on 328 newborns, performing a factor analysis on the behavioral items of the NBAS to provide reliability and validity data. Two main factors that appear to represent stable clusters across different socioeconomic, racial groups, geographic locations, and examiners are responsivity and reactivity, suggesting that items correlate highly with one another and are stable across groups. The items appear to exhibit internal consistency reliability.

Asch, Gleser, and Steichen (1986) examined the variability of certain observed cluster scores and the relative size of potential sources of measurement error, such as the examiner and an independent rater, and the occasion of testing. Appreciating sources of error of measurement is a concern in the construction of tests. Not to address measurement error is inconsistent with a primary AERA/APA/NCME standard (2.3), which specifically states that

> Each method of estimating a reliability that is reported should be defined clearly and expressed in terms of variance components, correlation coefficients, standard errors of measurement, percentages of correct decisions, or equivalent statistics. The conditions under which the reliability estimate was obtained and the situations to which it may be applicable should also be examined clearly. (AERA, APA, & NCME, 1985, p. 20)

Contributions to errors of measurement when reporting reliabilities in the manual are not delineated, although the goal in any reliability study reported is to estimate the magnitude of error in the scores obtained. Asch et al. (1986) conclude that when observed scores are averaged, the size of measurement error diminishes. For most of the clusters on the NBAS, three occasions with two raters per occasion are necessary to obtain minimum acceptable generalizability. In the 1984 manual, Lester (1984) urges the use of profile on recovery curves, which is a method for computing patterns of change over three NBAS exams. The procedure (outlined in the manual) is in line with the Asch et al. (1986) recommendations for averaging over scores and thereby reducing measurement error. The language utilized in the manual when discussing the method, however, is technical and may make utilizing the proposed formula difficult.

Training on the NBAS focuses on achieving examiner reliability in scoring and administration. Two observers are required to achieve an interrater agreement level of 90% by procuring the same performance and scoring it within 1 point of each other. The manual has explicit scoring criteria that are well defined. The degree to which two observers can achieve consistency in scoring is a measure of reliability when examining this scale. Als et al. (1977) stated that they could train individuals to produce reliability of 85% after 2 days. In this regard, reliability of the NBAS has been established, as all of the studies reviewed have reported an interobserver reliability coefficient of at least .85, which meet Nunnally's (1978) standards for basic research.

Problems arise, however, upon closer examination of the process of achieving the reliability criterion—that is, by reaching 100% agreement on 20 elicited reflex

items and a difference of 2 points or less on 2 of the 28 observed behavioral items on the scale. Dipietro and Larson (1989) caution, however, that an illusion of reliability exists because reliability is based on interobserver scoring methods and not on elicitation of performance. These investigators examined a sample of 100 neonates and found consistent examiner differences although interrater agreement was maintained throughout the testing. They suggest that the behavioral style of the examiner has an effect on the irritable responding of the neonate during the interactive portion of the exam. It may interfere with longitudinal predictability in neonatal development as well. Kaye (1978) states that the agreement between examiners concerning a specific observed behavior tells us nothing about the adequacy of the scale for assessing performance as a reliable characteristic of neonates. Perhaps there are underlying reactions of testers that cause them to score neonates in a similar way. Sameroff (1978) believes that handling techniques must be monitored during the examiner training period so that the level of stimulation a newborn receives is the same across examiners. The significance of the interrater reliability of the scales is questionable; raters scoring differently by 1 score point may still be considered reliable when in fact there may be error variance.

Lester (1984) utilized repeated NBAS cluster scores with two preterm infants at 3, 5, and 7 weeks after birth. They were followed up at 18 months of age and tested with the Bayley Scales of Infant Development (BSID), a well-standardized, reliable, and valid assessment of children's developmental status from 2 months to 2½ years. The profile curves computed from the NBAS were strongly related to 18-month outcome scores on orientation, motor, and two state clusters in a multiple regression analysis. All predictor scores were statistically significant, ranging from 0.42 to 0.63.

Sostik and Anders (1977) conducted a study investigating whether the NBAS predicts 10-week performance on the BSID, and they examined the relationship between early temperament and performance on both scales. Their study utilized 18 full-term infants. The results indicated that the prior Brazelton scoring and neonatal state control were predictive of Bayley mental quotients at 10 weeks. NBAS-identified temperamental intensity and distractibility at 2 weeks correlated later with the BSID 10-week results. However, the Brazelton dimensions did not relate to Bayley motor scale performance at all.

Walters, Vaughn, and Egeland (1980) demonstrated the predictive validity of the NBAS components, especially for the neonate's orienting ability. One hundred neonates were administered the NBAS at 7 and 10 days after birth and subsequently observed in the Ainsworth and Wittig (1969) strange-situation procedure at age 1. Results showed that infants classified as anxiously attached/resistant presented signs of unresponsiveness and motor immaturity at age 1.

The only long term follow-up study to date (Tronick & Brazelton, 1975) compares the predictive value of the NBAS with a standard neurological exam developed in the nationwide collaborative study sponsored by the National Institute for Nervous Disease and Stroke. The investigators studied 53 neonates who were deemed neurologically suspect by the standard neurological assessment. The NBAS was administered twice and a behavioral prediction was made, based on the sum of scores from the two neonatal exams and on the curve of recovery

implied by the results. The infants were followed at 4, 8, and 12 months, and again at 2, 4, 6, and 7 years of age. The predictive validity of the respective exams was measured against the 7-year outcome. The NBAS achieved a hit rate of 12 out of 15 children detected as suspect for neurological abnormalities, without including as many normal newborns in the suspect category as did the neurological exam (30 out of 38). The NBAS results indicated a lower false-alarm rate, while both exams were comparable in detecting children who were suspect for later neurological abnormalities. Although there appears to be a correlation between early and later performance, the results of the predictive validity studies should be viewed with caution in light of the small sample sizes used.

Critique

Since its conception 20 years ago, the Neonatal Behavioral Assessment Scale remains one of the most widely used scales in research. Brazelton and his colleagues' original aim, to offer a view of the healthy newborn seen in a social context, based on the way the neonate engages the caregiving environment and regulates internal control, has been achieved. The importance of assessing an infant's behavioral repertoire as early as possible through the changing matrices of states of consciousness has been established and met with success. Indeed, early intervention with the NBAS offers a strength that sets this instrument apart from other neonatal scales. The NBAS remains an excellent vehicle to enhance communication between caregiver, newborn, and researcher, as discussions about the newborn's ability begin while the exam is in process. By modeling constructive ways of interacting with the newborn during the exam, the assessment process promotes early therapeutic interventions. Although the available studies have methodologic flaws that prohibit drawing firm conclusions about positive effects of NBAS-based interventions, we know that parents are learning about their newborn's communication cues at the earliest level. Clinician and parent can begin to discuss their beliefs about the newborn's social availability and organization because they are uncovering these signals together, interpreting changes in skin color, sleeping, and crying behavior. The relative ease with which the NBAS can be administered, the explicit scoring criteria in the manual, and the simple testing kit make it a desirable instrument. However, the requisite training to achieve reliable administration and scoring is costly and time-consuming for the average clinician. Because multiple evaluation sessions are recommended over the first month as opposed to a single session, cost and time constraints are apparent.

Although the NBAS was never intended to be used as a test, it has been critiqued here as one because of the large volume of research utilizing it in this way, and as such its psychometric properties fall short of today's standards. There is a lack of normative data on which to compare the neonate's functioning. Standardization studies should be conducted, making it possible to compare a newborn's scale score with that of the general population of newborns and rendering the interpretation of the score meaningful based on stratified samples. The lack of independence between clusters reported in the studies reviewed is a related diffi-

culty that will have to be addressed. As set forth in *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985),

> Tests and testing programs should be developed on a sound scientific basis. Test developers should compile the evidence bearing on a test, decide which information is needed prior to test publication or distribution and which information can be provided later, and conduct any needed research. (Standard 3.5, p. 25)

If we are to view the NBAS as a test, then it does not comply with this primary standard. Basic research concerning normative data should have been conducted prior to publication of the manual.

Although predictive validity studies have met with some success based on small sample sizes, few long-term validation studies have been completed. Additional long-term validation studies should be incorporated into the manual to support its scientific claims. However, despite limited evidence of validity, the NBAS does comply with Standard 1.11 regarding the descriptions provided of subject samples and the statistical analyses used to determine the degree of predictive accuracy. Another primary standard (1.2) the test developers of the NBAS have complied with states that "if validity for some common interpretation has not been investigated, the fact should be made clear, and potential users should be cautioned about making such interpretations" (AERA, APA, & NCME, 1985, p. 13). Horowitz and Linn (1984) caution that significant correlations reported account for small measures of outcome variance because environmental variables have not been included in combination with NBAS measures. However, subsequent studies have demonstrated predictive validity when correlating NBAS scores with later scores on the BSID. Thus, one should view the evidence for predictive validity cautiously at this time.

The test-retest designs reported to assess the NBAS are inappropriate. In the manual, Lester (1984) questions the value of test-retest reliability on the NBAS using a Pearson *r* strategy. Another means of assessing reliability therefore should be considered, such as measures of internal consistency. AERA/APA/NCME Standard 2.6 specifies that "coefficients based on internal analysis should not be interpreted as substitutes for alternate-form reliability or estimates of stability over time unless other evidence supports that interpretation in a particular context" (1985, p. 21), and the NBAS does not comply. There appears to be insufficient evidence supporting test-retest reliability of the NBAS, and its validity needs further substantiation as well by long-term prediction studies. Certainly there has been evidence to justify substituting internal consistency reliability over test-retest reliability in the literature reported (Nunnally, 1978).

Because newborn behavior changes rapidly over the first month of life, it is not likely that high test-retest reliability will occur for neonatal assessments such as the NBAS. The measures of interrater reliability are high, but Clopton and Martin (1985) have suggested that subtle differences in examiner style as well as the 1-point difference in scoring may introduce substantial error variance into scoring the scale. Despite the large body of research focused on the NBAS, conclusions about its reliability cannot be drawn at this time.

The NBAS appears to meet acceptable standards for basic research at this time

(Nunnally, 1978), although it is not a clinically tested tool, as Brazelton will confirm. It does not proport to be a well-standardized instrument, and caution must be exercised when administering the scale in clinical practice as well as in research. Those using the NBAS would benefit from a comprehensive normative data study, long-term predictive validity data, and additional reliability studies based on the internal consistency reliability model described in Nunnally (1978). Generalizations to the population at large cannot be made until more appropriate test designs are established. If refined in terms of its psychometric properties, the NBAS might be used to define individual characteristics in a more meaningful way.

References

This list includes text citations and suggested additional reading.

Ainsworth, M., & Wittig, R. (1969). Attachment and exploratory behavior of one year olds in a strange situation. In B. Fuss (Ed.), *Determinants of infant behavior.* New York: Barnes & Noble.

Aleksandrowiez, M.K., & Aleksandrowiez, O.R. (1974). The molding of personality. *Child Psychiatry and Human Development, 5*(4), 231–241.

Als, H. (1978). Assessing an assessment: Conceptual considerations, methodological issues and a perspective on the future of the Neonatal Behavioral Assessment Scale. In A.J. Sameroff (Ed.), *Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale* (Monographs of the Society for Research in Child Development, 43[5–6, Serial No. 177], pp. 14–28).

Als, H. (1984). Newborn Behavioral Assessment. In W. Burns & J. Lavigne (Eds.), *Progress in pediatric psychology* (pp. 1–46). Orlando, FL: Grune & Stratton.

Als, H. (1986). A synactive model of neonatal behavioral organization: Framework for the assessment of neurobehavioral development in the premature infant and for support of infants and parents in the neonatal intensive care environment. *Physical and Occupational Therapy in Pediatrics, 6*(3–4), 3–53.

Als, H., Lester, B., Tronick, E., & Brazelton, B. (1982). Manual for the Assessment of Preterm Infants' Behavior (APIB). In H.E. Fitzgerald, B.M. Lester, & M.W. Yogman (Eds.), *Theory and research in behavioral pediatrics* (Vol. 1). New York: Plenum.

Als, H., Tronick, E., Adamson, L., & Brazelton, T.B. (1976). The behavior of the full-term but underweight newborn infant. *Developmental Medicine and Child Neurology, 18,* 590–602.

Als, H., Tronick, E., Lester, B.M., & Brazelton, T.B. (1977). The Brazelton Neonatal Behavioral Assessment Scale (BNBAS). *Journal of Abnormal Child Psychology, 5*(3), 215–229.

Als, H., Tronick, E., Lester, B.M., & Brazelton, T.B. (1979). The Brazelton Neonatal Scale (NBAS). In H.J. Osofsky (Ed.), *Handbook of infant development.* New York: Wiley.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Anderson, C.J., & Sawin, D.B. (1983). Enhancing responsiveness in mother-infant interaction. *Infant Behavior and Development, 6,* 361–368.

Anderson, L.T., Coll, C.G., Vohr, B.R., Emmons, L., Brann, B., Shaul, P.W., Mayfield, S.R., & Oh, W. (1989). Behavioral characteristics and early temperament of premature infants with intracranial hemorrhage. *Early Human Development, 18,* 273–283.

Asch, P.A., Gleser, G.C., & Steichen, J. (1986). Dependability of Brazelton Neonatal Behavioral Assessment cluster scales. *Infant Behavior and Development, 9,* 291–306.

Aylward, G.P., Lazzara, P., & Meyer, A. (1978). Behavior and neurological characteristics of a hydranencephalic infant. *Developmental Medicine and Child Neurology, 20,* 211–217.

Beal, J.A. (1986). The Brazelton Neonatal Behavioral Assessment Scale: A tool to enhance parental attachment. *Journal of Pediatric Nursing, 1*(3), 170–177.

Beal, J.A. (1989). The effect on father-infant interactions of demonstrating the Neonatal Behavioral Assessment Scale. *Birth, 16*(1), 18–22.

Belsky, J. (1985). Experimenting with the family in the newborn period. *Child Development, 56,* 407–414.

Brazelton, T.B. (Ed.). (1973). Neonatal Behavioral Assessment Scale [Special issue]. *Clinics in Developmental Medicine, 50.*

Brazelton, T.B. (1978). The remarkable talents of the newborn. *Birth and the Family Journal, 5*(4), 187–191.

Brazelton, T.B. (Ed.). (1984). *Neonatal Behavioral Assessment Scale* (2nd ed.). London: Spastics International Medical Publications.

Brazelton, T.B., & Freedman, D.G. (1971). Manual to accompany Cambridge Newborn Behavior and Mental Scales. In G.B.A. Stoelings & J.J. Van Der Werff Ten Bosch (Eds.), *Normal and abnormal development of brain and behavior.* Leiden, The Netherlands: Leiden University Press.

Brazelton, T.B., Koslowski, B., & Tronick, E., (1971). Neonatal behavior among urban Zombians and Americans. *Journal of the American Academy of Child Psychiatry, 15*(1), 97–107.

Brazelton, T.B., Tronick, E., Lechtig, A., Lasky, R.E., & Klein, R.E. (1977). The behavior of nutritionally deprived Guatemalan infants. *Developmental Medicine and Child Neurology, 19,* 364–372.

Burns, W.J., Deddish, R.B., & Hatcher, R.P. (1982). Developmental assessment of premature infants. *Developmental and Behavioral Pediatrics, 3*(1), 12–17.

Chasnoff, I.J., Burns, K.A., & Burns, W.J. (1985). Cocaine use in pregnancy: Perinatal morbidity and mortality. *Neurotoxicology and Teratology, 9,* 291–293.

Chasnoff, I.J., Burns, W.J., Hatcher, R.P., & Burns, K.A. (1983). Phencyclidine: Effects on the fetus and neonate. *Developmental Pharmacology Therapeutics, 6,* 404–408.

Chasnoff, I.J., Burns, W.J., & Schnoll, S. (1983). Perinatal addiction: The effects of maternal narcotic and non-narcotic substance abuse on fetus and neonate. *National Institute on Drug Abuse Research Monograph Series, 49,* 220–226.

Chasnoff, I.J., Burns, W.J., Schnoll, S.H., & Burns, K.A. (1985). Cocaine use in pregnancy. *The New England Journal of Medicine, 313*(11), 666–669.

Chasnoff, I.J., Griffith, D.R., MacGregor, S., Dirkes, K., & Burns, K.A. (1989). Temporal patterns of cocaine use in pregnancy. *Journal of the American Medical Association, 261*(12), 1741–1744.

Chasnoff, I.J., Hatcher, R., & Burns, W.J. (1982). Polydrug and methadone-addicted newborns: A continuum of impairment. *Pediatrics, 70*(2), 210–273.

Chasnoff, I.J., Hatcher, R., Burns, W.J., & Schnoll, S. (1983). Pentazocine and tripelennamine (t's and blue's): Effects on the fetus and neonate. *Developmental Pharmacology Therapeutics, 6,* 162–169.

Chasnoff, I.J., Schnoll, S.H., Burns, W.J., & Burns, K.A. (1984). Maternal neonarcotic substance abuse during pregnancy: Effects on infant development. *Neurobehavioral Toxicology and Teratology, 6,* 277–280.

Choi, E.S., & Hamilton, R.K. (1986). The effects of culture on mother-infant interaction. *Journal of Obstetric, Gynecologic and Neonatal Nursing, 15*(3), 256–261.

Claire, K.L. (1978). Neonatal assessment procedures: A historical review. *Child Development, 49,* 5–26.

Clopton, N., & Martin, A.S. (1985). A criticism of interrater reliability procedures for the Brazelton Neonatal Behavioral Assessment Scale. *Physical and Occupational Therapy in Pediatrics, 4*(4), 55–65.

Coles, C.D., Smith, I.E., & Falek, A. (1987). Prenatal alcohol exposure and infant behavior: Immediate effects and implications for later development. In *Children of alcoholics* (pp. 87–105). Binghamton, NY: Haworth.

Coles, C.D., Smith, I.E., Fernhoff, P.N., & Falek, A. (1985). Neonatal neurobehavioral characteristics as correlates of maternal alcohol use during gestation. *Alcoholism: Clinical and Experimental Research, 9*(5), 454–460.

Coles, C.D., Smith, I.E., Lancaster, J.S., & Falek, A. (1987). Persistence over the first month of neurobehavioral differences in infants exposed to alcohol prenatally. *Infant Behavior and Development, 10*, 23–37.

Crockenberg, S., & Acredolo, C. (1983). Infant temperament ratings: A function of infants, mothers, or both? *Infant Behavior and Development, 6*, 61–72.

Crockenberg, S., & McClusky, K. (1986). Change in maternal behavior during the baby's first year of life. *Child Development, 57*, 746–753.

Crockenberg, S.B. (1981). Infant irritability, mother responsiveness, and social support: Influences on the security of infant-mother attachment. *Child Development, 52*, 857–865.

DeVries, M., & Super, C.M. (1978). Contextual influences on the Neonatal Behavioral Assessment Scale and implications for its cross-cultural use. In A. Sameroff (Ed.), *Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale* (Monographs of the Society for Research in Child Development, *43*[5–6, Serial No. 177], pp. 92–100).

DiPietro, J.A., & Larson, S.K. (1989). Examiner effects in the administration of the NBAS: The illusion of reliability. *Infant Behavior and Development, 12*, 119–123.

Dunlop, K.H. (1982). Neonatal behavior in an urban working-class Mexican population: Relationships with drug condition and parity. *Interamerican Journal of Psychology, 15*(2), 79–95.

Escher-Graub, D.C., & Fricker, H.S. (1986). Jaundice and behavioral organization in the full-term neonate. *Helvetica Paediatrica Acta, 41*, 425–435.

Emory, E.K., Konopka, S., Hronsky, S., Tuggey, R., & Dave, R. (1988). Salivary caffeine and neonatal behavior. *Psychopharmacology, 94*, 64–68.

Emory, E.K., Tynam, W.D., & Dave, R. (1989). Neurobehavioral anomalies in neonates with seizures. *Journal of Clinical and Experimental Neuropsychology, 11*(2), 231–240.

Ernhart, C.B., Wolf, A.W., Linn, P.L., Sokoi, R.J., Kennard, M.J., & Filipovich, H.F. (1985). Alcohol-related birth defects: Syndromal anomalies, intrauterine growth retardation, and neonatal behavioral assessment. *Alcoholism: Clinical and Experimental Research, 9*(5), 446–453.

Field, T. (1979). Visual and cardiac responses to animate and inanimate faces by young term and preterm infants. *Child Development, 50*, 188–194.

Finnegan, L.P. (1984, April). *Effects of maternal opiate abuse on the newborn.* Paper presented at the 68th Annual Meeting of the Federation of American Societies for Experimental Biology, St. Louis, MO.

Fitzgerald, H.E., Lester, B.M., & Yooman, M.W. (Eds.). (1982). *Theory and research in behavioral pediatrics: Vol. 1.* New York: Plenum.

Frances, P.L., Self, P.A., & McCaffree, M. (1984). Behavioral assessment of a hydranencephalic neonate. *Child Development, 55*, 262–266.

Gibbs, R.M. (1981). *Clinical uses of the Brazelton Neonatal Behavioral Assessment Scale in nursing practice* (Adapted from a presentation sponsored by the Johnson & Johnson Baby Products Co. and given at the NAPNAP First Annual Nursing Conference on Pediatric Primary Care, Feb. 27–Mar. 1, 1980, Washington, DC). Washington, DC: NAPNAP.

Golas, G.A., & Parks, P. (1986). Effects of early postpartum teaching on primiparas: Knowledge of infant behavior and degree of confidence. *Research in Nursing and Health, 9*, 209–214.

Gomes-Pedro, J., Bento de Almedia, Silveria da Costa, C., & Barbosa, A. (1984). Influence of

early mother-infant contact on dyadic behavior during the first month of life. *Developmental Medicine and Child Neurology, 26,* 657–664.

Gottlieb, S., & Barrett, D.E. (1986). Effects of unanticipated cesarean section on mothers, infants, and their interaction in the first month of life. *Developmental and Behavioral Pediatrics, 7*(3), 180–185.

Greene, J.G., Fox, N.A., & Lewis, M. (1983). The relationship between neonatal characteristics and three-month mother-infant interaction in high-risk infants. *Child Development,* 54, 1286–1296.

Gyurke, J.S. (1988). An examination of the effectiveness of multiple summary scoring procedures of the BNBAS in detecting group differences. *Infant Mental Health Journal, 9*(3), 201–208.

Hofheimer, J.A., Poisson, S.S., Strauss, M.E., Eyler, F.D., & Greenspan, S.I. (1983). Perinatal and behavioral characteristics of neonates born to multi-risk families. *Developmental and Behavioral Pediatrics, 4*(3), 163–170.

Horowitz, F.D., & Brazelton, T.B. (1973). Research with the Brazelton neonatal scale. In T.B. Brazelton (Ed.), Neonatal Behavioral Assessment Scale [Special issue], *Clinics in Developmental Medicine, 50.*

Horowitz, F.D., & Linn, P.L. (1984). Use of the NBAS in research. In T.B. Brazelton (Ed.), *Neonatal Behavioral Assessment Scale* (2nd ed.). London: Spastics International Medical Publications.

Horowitz, F.D., Sullivan, J.W., & Linn, P. (1978). Stability and instability in the newborn infant: The quest for elusive threads. In A.J. Sameroff (Ed.), *Organization and stability of newborn behavior: A commentary on the Brazelton Neonatal Behavioral Assessment Scale* (Monographs of the Society for Research in Child Development, 43[5–6, Serial No. 177], pp. 29–45).

Jacobson, J.L., Jacobson, S.W., Fein, G.G., & Schwartz, P.M. (1984). Factors and clusters for the Brazelton scale: An investigation of the dimensions of neonatal behavior. *Developmental Psychology, 20*(3), 339–353.

Jeremy, R.J., & Hans, S. (1985). Behavior of neonates exposed in utero to methadone as assessed on the Brazelton scale. *Infant Research and Development, 8,* 323–336.

Jones, L.C., & Lenz, E.R. (1985). Father-newborn interaction: Effects of social competence and infant state. *Nursing Research, 35*(3), 149–153.

Kay, K. (1978). Discriminating among normal infants by multivariate analyses of Brazelton scores. *Monographs of the Society for Research in Child Development, 43*(5–6).

Keefer, C.H., Tronick, E., Dixon, S., & Brazelton, T.B. (1982). Specific differences in motor performance between Gusii and American newborns and a moderation of the Neonatal Behavioral Assessment Scale. *Child Development, 53,* 754–759.

Kestermann, G. (1981). Assessment of individual differences among healthy newborns on the Brazelton scale. *Early Human Development, 5,* 15–27.

Kline, P. (1986). *A handbook of test construction.* New York: Methuen.

Kochanevich-Wallace, P.M., McCluskey-Faucett, K.A., & Meek, N.E. (1987). Method of delivery and parent-newborn interaction. *Journal of Pediatric Psychology, 13*(2), 213–221.

Kron, R.E., Kaplan, S.L., Finnegan, L.P., Litt, M., & Phoenix, M.D. (1975). The assessment of behavioral change in infants undergoing narcotic withdrawal: Comparative data from clinical and objective methods. *Addictive Diseases: An International Journal, 2*(2), 257–275.

Kuhnert, B.R., Linn, P.L., Kennerd, M.J., & Kuhnert, P.M. (1985). Effects of low doses of meperidine on neonatal behavior. *Annals of Analgesia, 64,* 335–342.

Lancioni, G.E., Horowitz, F.D., & Sullivan, J.W. (1980). The NBAS-K 1: A study of its stability and structure over the first month of life. *Infant Behavior and Development, 3,* 341–359.

Leijon, I. (1982). Assessment of behavior on the Brazelton scale in healthy preterm infants from 32 conceptional weeks until full-term age. *Early Human Development, 7,* 109–118.

Leijon, I., Finnstrom, D., Hedenskog, S., Ryden, G., & Tylleskar, J. (1979). Spontaneous labor

and elective induction—A prospective randomized study. *Acta Paediatrica Scandinavica, 68,* 553–560.

Lester, B.M., (1984). Data analyses and prediction. In T.B. Brazelton (Ed.), *Neonatal Behavioral Assessment Scale* (2nd ed.). London: Spastics International Medical Publications.

Lester, B.M., Als, H., & Brazelton, T.B. (1982). Regional obstetric anesthesia and newborn behavior: A reanalysis towards synergistic effects. *Child Development, 53,* 687–692.

Lester, B.M., Garcia-Coll, C.G., & Sepkoski, C. (1982). Teenage pregnancy and neonatal behavior: Effects in Puerto Rico and Florida. *Journal of Youth and Adolescence, 11*(5), 385–402.

Lester, B.M., Garcia-Coll, C.T., Valcaicel, M., Hoffman, J., & Brazelton, T.B. (1986). Effects of atypical patterns of fetal growth on newborn (NBAS) behavior. *Child Development, 57,* 11–19.

Maloni, J.A., Stegman, C.E., Taylor, P.M., & Brownell, C.A. (1986). Validation of infant behavior identified by neonatal nurses. *Nursing Research, 35*(3), 133–136.

Maier, R.A., Jr., Prinn, S.M., Nagy, J.N., Halmes, D.L., & Slaymaker, F. (1983). A methodological note on the use of a priori cluster scores for the Brazelton Neonatal Behavioral Assessment Scale (BNBAS). *Infant Behavior and Development, 6,* 299–303.

Meyers, B.J. (1982). Early intervention using Brazelton training with middle-class mothers and fathers of newborns. *Child Development, 53,* 462–471.

Muhlen, L., Pryke, M., & Wade, K. (1986). Effects of type of birth anaesthetic on Neonatal Behavioral Assessment Scale scores. *Australian Psychologist, 21,* 253–271.

Murphy, T.F., Scher, M.S., Klesh, K.W., & Guthrie, R.D. (1988). Easy neurobehavioral abnormalities in infants with persistent pulmonary hypertension of the newborn. *Infant Behavior and Development, 11,* 159–167.

Nugent, K.J. (1980, February/March). *The Brazelton Neonatal Behavioral Assessment Scale: Implications for intervention.* Paper presented at the NAPNAP First Annual Nursing Conference on Pediatric Primary Care, Washington, DC.

Nugent, K.J., & Brazelton, T.B., (1989). Preventive intervention with infants and families: The NBAS model. *Infant Mental Health Journal, 10*(2), 84–99.

Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Osofsky, H.J. (1974). Relationships between prenatal medical and nutritional measures, pregnancy outcome, and early infant development in an urban poverty setting. *American Journal of Obstetrics and Gynecology, 123*(7), 682–690.

Osofsky, J.D. (1976). Neonatal characteristics and mother-infant interaction in two observational studies. *Child Development, 47,* 1138–1147.

Osofsky, J.D., & Danzger, B. (1974). Relationships between neonatal characteristics and mother-infant interaction. *Developmental Psychology, 10*(1), 124–130.

Osofsky, J.D., & O'Connell, E.J. (1977). Patterning of newborn behavior in an urban population. *Child Development, 48,* 532–536.

Paludetto, R., Mansi, G., Rinaldi, P., DeLuca, T., Corchia, C., DeCurtis, M., & Andolfi, M. (1982). Behavior of preterm newborns reaching term without any serious disorder. *Early Human Development, 6,* 357–363.

Paludetto, R., Rinaldi, P., Mansi, G., Andolfi, M., & Del Giudice, G. (1984). Early behavioral development of preterm infants. *Developmental Medicine and Child Neurology, 26,* 347–352.

Penman, R., Meares, R., Baker, K., & Milgrom-Freedman, J. (1983). Synchrony in mother-infant interaction: A possible neuro-physiological base. *British Journal of Medical Psychology, 56,* 1–7.

Rogan, W.J., Gladen, B.C., McKinney, J.D., Carreras, N., Hardy, P., Thullen, J., Tinglestad, J., & Tully, M. (1986). Neonatal effects of transplacental exposure to PCB's and DDE. *Journal of Pediatrics, 109,* 335–341.

Saco-Pollitt, C. (1981). Birth in the Peruvian Andes: Physical and behavioral consequences in the neonate. *Child Development, 52,* 839–846.

Sameroff, A.J. (Ed). (1978). Organization and stability of newborn behavior: A commentary

on the Brazelton Neonatal Behavioral Assessment Scale. *Monographs of the Society for Research in Child Development, 43*(5–6, Serial No. 177).

Saxton, D.W. (1978). The behavior of infants whose mothers smoke in pregnancy. *Early Human Development, 44*, 363–369.

Scafidi, F.A., Field, T.M., Schanberg, S.M., Bauer, C.R., Vega-Lahr, N., Garcia, R., Poirer, J., Nystrom, G., & Kuhn, C. (1986). Effects of tactile/kinesthetic stimulation on the clinical course and sleep/wake behavior of preterm neonates. *Infant Behavior and Development, 9*, 91–105.

Scanlon, K.B., Scanlon, J.W., & Tronick, E. (1984). The impact of prenatal and neonatal events on the early behavior of the extremely premature human. *Developmental and Behavioral Pediatrics, 5*(2), 65–73.

Silverman, C.R. (1991). *The Neonatal Behavioral Assessment Scale (NBAS): A review and critique.* Unpublished doctoral dissertation, Nova University, Ft. Lauderdale, FL.

Smith, I.E., Coles, C.D., Lancaster, J., Fernhoff, P.M., & Falek, A. (1986). The effect of volume and duration of prenatal exposure on neonatal physical and behavioral development. *Neurobehavioral Toxicology and Teratology, 8*, 375–381.

Solkoff, N., & Matuszak, D. (1975). Tactile stimulation and behavioral development among low-birthweight infants. *Child Psychiatry and Human Development, 6*(1), 33–37.

Sostik, A.M., & Anders, T.F. (1977). Relationships among the Brazelton neonatal scale, Bayley infant scales and early temperament. *Child Development, 48*, 320–323.

Soule, B.A., Standley, K., Copans, S.A., & Davis, M. (1974). Clinical uses of the Brazelton Scales. *Pediatrics, 54*(5), 583–586.

St. Clair, K.L. (1978). Neonatal assessment procedures: A historical review. *Child Development, 49*, 280–292.

Standley, K.A., Soule, B.A., Copans, S.A., & Duchouny, A. (1974). Local regional anesthesia during childbirth: Effect on newborn behaviors. *Science, 186*, 634–635.

Standley, K.A., Soule, B.A., Copans, S.A., & Klein, R.P. (1978). Multidimensional sources of infant temperament. *Genetic Psychology Monographs, 98*, 203–231.

Strauss, M.E., Lessen-Firestone, J.K., Starr, R.H., Jr., & Ostera, E.M., Jr. (1975). Behavior of narcotics-addicted newborns. *Child Development, 46*, 887–893.

Strauss, M.E., Starr, R.H., Jr., Ostera, E.M., Jr., Chavez, C.J., & Stryker, J.C. (1976). Behavioral concomitants of prenatal addiction to narcotics. *Journal of Pediatrics, 89*(5), 842–846.

Streissguth, A.P., Barr, H.M., & Martin, D.C. (1983). Maternal alcohol use and neonatal habituation assessed with the Brazelton scale. *Child Development, 54*, 1109–1118.

Szajnberg, N., Ward, M.J., Krauss, A., & Kessler, D.B. (1986). Low birth weight prematures: Preventative intervention and maternal attitude. *Child Psychiatry and Human Development, 17*, 152–165.

Tronick, E., & Brazelton, T.B. (1975). Clinical uses of the Brazelton neonatal assessment. In B.Z. Friedlander, G.M. Sterritt, & G.E. Kirk (Eds.), *Exceptional infant 3: Assessment and intervention*. New York: Brunner/Mazel.

Tronick, E.Z. (1987). The Neonatal Behavioral Assessment Scale as a bio-marker of the effects of environmental agents on the newborn. *Environmental Health Perspectives, 74*, 185–189.

Vaughn, J., Brown, J., & Carter, J.P. (1986). The effects of maternal anemia on infant behavior. *Journal of the National Medical Association, 78*(10), 963–968.

Walters, E., Vaugh, B.E., & Egeland, B.R. (1980). Individual differences in infant-mother attachment relationships at age one: Antecedents in neonatal behavior in an urban, economically disadvantaged sample. *Child Development, 51*, 208–216.

Widmayer, S.M., & Field, T.M. (1980). Effects of Brazelton demonstrations on early interactions of preterm infants and their teenage mothers. *Infant Behavior and Development, 3*, 79–89.

Widmayer, S.M., & Field, T.M. (1981). Effects of Brazelton demonstrations for mothers on the development of preterm infants. *Pediatrics, 67*(5), 711–714.

Worobey, J. (1986a). Convergence among assessments of temperament in the first month. *Child Development, 57,* 47–55.

Worobey, J. (1986b). Neonatal stability and one-month behavior. *Infant Behavior and Development, 9,* 119–124.

Worobey, J., & Belsky, J. (1982). Employing the Brazelton scale to influence mothering: An experimental comparison of three strategies. *Developmental Psychology, 18*(5), 736–743.

Worobey, J.W., & Brazelton, T.B. (1986). Experimenting with the family in the newborn period: A commentary. *Child Development, 57,* 1298–1300.