Published in final edited form as:

Infant Behav Dev. 2020 February ; 58: 101408. doi:10.1016/j.infbeh.2019.101408.

Infant Behaviors and Maternal Parenting Practices: Short-Term Reliability Assessments

Marc H. Bornstein^{1,2}, Chun-Shin Hahn¹, Diane L. Putnick¹, Gianluca Esposito^{3,4}

¹*Eunice Kennedy Shriver National Institute of Child Health and Human Development*

²*Institute for Fiscal Studies, London, UK*

³*Nanyang Technological University, Singapore*

⁴*University of Trento, Italy*

Abstract

Consistency in the order of individuals in a group across short periods of time—*reliability*—is both important developmentally and meaningful psychologically. For example, documenting the reliabilities of infant behaviors and maternal parenting practices elucidates the nature and structure of early development. In this prospective short-term longitudinal study ($Ns = 51$ 5-month infants and their mothers), we examined reliabilities of individual variation in multiple infant behaviors (physical development, social interaction, exploration, nondistress vocalization, and distress communication) and maternal parenting practices (nurturing, encouragement of motor growth, social exchange, didactic interaction, provision of the material environment, and speech to infant). Medium to large effect size reliabilities characterize infant behaviors and maternal parenting practices, but both betray substantial amounts of unshared variance. Established reliability is essential to the application of these measures in infancy studies, it is central to replication, and it is a limiting factor in predictive validity.

Keywords

infancy; parenting; reliability; replication; validity

Within any group and at every age, human beings normally vary (sometimes dramatically) amongst themselves on any given characteristic, and that variation usually appears as a normal (Gaussian) distribution in the population. A core and recurring issue in developmental and psychological science is the *reliability* – that is, individual-order short-

Address correspondence to: Dr. Marc H. Bornstein, *Eunice Kennedy Shriver National Institute of Child Health and Human Development*, 8404 Irvington Avenue, Bethesda, MD 20817 USA, TEL: 301-656-1642, marc.h.bornstein@gmail.com.
CRediT author statement

Marc H. Bornstein: Conceptualization, Methodology, Funding acquisition, Writing- Original draft preparation, Supervision, Project administration, Resources, Methodology, Writing- Reviewing and Editing; Chun-Shin Hahn: Data curation, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing; Diane L. Putnick: Data curation, Formal analysis, Writing- Original draft preparation, Writing- Reviewing and Editing; Gianluca Esposito: Writing- Reviewing and Editing.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

term consistency through time – of that variation (see Anastasi, 1968; Baltes & Nesselroade, 1979; Bornstein, Putnick, & Esposito, 2017; Cairns, 1979; Cohen, 1988; DeVellis, 2016; Hartmann, Abbott, & Pelzel, 2015; McCall, 1981; Miller, 1987; Nunnally, 2017; Wohlwill, 1973; Yarrow & Waxler, 1979). Operationally, a reliable (temporally stable, interindividually consistent) characteristic is one that some individuals display at relatively high levels at one point in time and again display at relatively high levels at a second point a short time later, where other individuals display lower levels at both times; an unreliable characteristic is one where individuals do not maintain relative order in their group even across a short duration. The prototypical model describes *homotypic reliability*, the maintenance of order among individuals on the same characteristic across a short period of time. This study documents short-term homotypic reliabilities in infant behaviors and maternal parenting practices in the first year of life.

Meaningfulness of Developmental Reliability

“Reliability is a fundamental issue in psychological measurement. Its importance is clear once its meaning is fully understood” (DeVellis, 2016, p. 27); indeed, some authors have pointed to the significance of reliability as “a badge for quality of the data” (Christa, Seidl, Singh, & Houston, 2016, p. 661). The psychometric study of reliability of psychological characteristics (whether constructs, structures, functions, or processes) in individuals is central for practical, theoretical, substantive, clinical, and methodological reasons. First, practically, many systems in humans require certain consistencies – physical, chemical, psychological, and environmental homeostases – to survive. These systems allow for the maintenance of consistencies even in the face of changing circumstances; thus, living organisms exist in, and strive to maintain, states of “adaptive” consistency (Cannon, 1932). Second, theoretically, reliability of parenting is often assumed as central to theories of family systems functioning. Radke-Yarrow, Zahn-Waxler, and Chapman (1983, pp. 501-502) posited that “in theories of childrearing, parental behavior is assumed to have effects on infants through a history of experiences. There is faith that, over time, parental influences lead to generalized behavioral tendencies that have some durability.” For example, attachment theory asserts consistency in neurobiological systems that underpin, and relational behaviors that express, affiliative bonds (Sroufe, Egeland, Carlson, & Collins, 2005). More broadly, reliability speaks to trait versus state, person versus situation, perspectives on psychological – hence child and parent – functioning (Fleeson, 2004). The trait/person position argues that dispositional characteristics drive individuals to act similarly at different times, whereas the state/situation position argues that immediate circumstances determine behavior so people act differently at different times (Epstein, 1979; Mischel, 1979). Third, substantively, reliability provides basic information about development as it is developmentally informative to describe an individual or a characteristic as reliable or not over time. Only relatively reliable characteristics would be expected to quantify meaningful differences between people. Whether infants or parents maintain their order in a group across short periods of time therefore informs not only about individual variation, but also contributes to understanding possible origins, nature, and future of those characteristics. Fourth, clinically, for measures to be incorporated into diagnostic batteries (to validly measure concurrent characteristics or predict future ones), performance at a given time needs

to be a reliable indicator of the individual (DeVellis, 2016). Finally, methodologically, reliability has multiple implications for measurement in developmental science. The reliability of a characteristic sets a specific statistical limit on that characteristic's long-term stability or predictive validity (Alder & Scher, 1994; Nunnally, 2017). If validity is indexed by a characteristic's correlation with a criterion (r_{xy}), and the characteristic's reliability is expressed as r_{xx} , the upper limit of r_{xy} is $r_{xx}^{1/2}$. The lower the reliability the less confidence in utilizing and interpreting the characteristic (Hartmann et al., 2015; Maloney & Ward, 1976). Focusing on reliability also constructively responds to the issue of replication in that reliability is a critical requirement for reproducible research. In brief, reliability has many basic applications and significant implications in developmental science.

Some Preliminary Design and Analytic Considerations

In designing a reliability study such as this, three preliminary temporal considerations must be addressed: the duration of the inter-assessment interval, the duration of the observation, and the age of the participants to be studied. Any reliability study has to fix on some parameter of these three temporal characteristics. The question is to reasonably justify each of the parameters selected. Here, we examined the behaviors of infants and the parenting practices of mothers at each of two times over a short (1-week) period for an optimal duration each time (1 hour) in young (5-month) babies.

Spacing of observations is the first temporal design consideration: Notably, “short-term reliability” and “long-term stability” lie on a temporal continuum with no clear or agreed-upon psychometric demarcation between the two; that is, how long an inter-assessment interval is that distinguishes reliability versus stability -- an hour? a day? a week? a month? a year? – is not a settled matter. If the interval is too short, potential “panel conditioning” problems of practice, memory, familiarity, and other carry-over effects must be contended with as these would tend to artificially increase the reliability estimate. Other impediments to conducting too closely spaced observations also present themselves, such as expense and logistical difficulties, unacceptable attrition, data character, and measurement itself may become a treatment. In consequence, closely spaced measurements could attenuate or maximize reliability assessment and, potentially, yield false estimates of reliability. By contrast, too widely spaced measurements might suffer similar biases on account of real changes in infants and parents over time that would attenuate reliability. Certainly, for fast-developing infants an inter-assessment interval of as short as a month, say from 5 to 6 months, would in actuality represent a change of approximately 20% of their life time. In consequence, a month would likely eventuate in spuriously low estimates of reliability. Temporal stability is consistently negatively correlated with the length of time between assessments (the Guttman simplex); indeed, the decreasing correlation with increasing inter-assessment time has been observed so commonly that it has assumed the character of a basic law of behavior. For these reasons we settled on a 1-week interobservation assessment interval (which matches several reliability designs already in the published literature reviewed below).

Duration of the observations is a second temporal design consideration: We studied 50 min of continuous behaviors of infants and parenting practices of mothers. Meta-analysis has

shown this duration to fall in an “optimal” recording time frame for mother-child interaction (see Holden & Miller, 1999, p. 239).

Age of the infants is a third temporal design consideration: We studied infants at 5 to 6 months of age. By the middle of the first year, the infant’s scope of apperception has broadened to the dyad and beyond; no longer *fetus ex utero*, infants are alert for extended periods of time, are becoming regulated in their emotions, increasingly initiate interactions using directed social behaviors like gaze, actively participate in reciprocal exchanges, and explore the environment visually and tactually (see Bornstein, Arterberry, & Lamb, 2014). This period in the middle of the first year is also relatively settled developmentally and follows and precedes transitionary phases, bio-behavioral shifts, psychic organizations, developmental crises, and the like (Erikson, 1963; Piaget, 1952; Spitz, 1965; Trevarthen, 1988) that would artificially disrupt reliability measurement. We designed this longitudinal study of the short-term homotypic reliabilities of multiple infant behaviors and maternal parenting practices with all these preliminary temporal issues about reliability in mind.

Finally, reliability is typically assessed by Pearson correlation coefficients. In describing effect sizes, we follow Cohen’s (1988, pp. 79-80) terminology; small effect size estimate of population correlation, $r = .10$; medium effect size, $r = .30$; large effect size, $r = .50$ (see also Landis & Koch, 1977; Weir, 2005).

Extant Studies of Reliability in Infants and Parenting Infants

In infants, we studied short-term homotypic reliabilities of multiple commonly used age-appropriate behaviors that are principal gauges of state of arousal as well as of cognitive, communicative, emotional, and social functioning. Among mothers, we examined short-term homotypic reliabilities of frequent and prominent parenting practices of infants. A plethora of reliability studies of parent reports of, for example, infant temperament populate the developmental literature (e.g., Medoff-Cooper, Carey, & McDevitt, 1993, reported large [$r = .43$ to $.87$] 2- to 3-week reliabilities of mothers’ reports of infant temperament on the Early Infancy Temperament Questionnaire for 404 1- to 4-month-old infants). However, for all the many reports about infancy and (maternal) parenting, surprisingly few studies of the first-order short-term reliability of actual infant *behaviors* and maternal parenting *practices* have been published. Introducing this literature here, we document those studies limiting our review to within-child across-time reports of short-term homotypic reliability of actual term infant behaviors and maternal parenting practices in the first year of life.

Reliability in Infant Behaviors.

Coates, Anderson, and Hartup (1972) reported the reliability of discrete attachment behaviors (visual regard, vocalizing, touching, and proximity seeking) in $Ns = 22$ -28 10-, 14-, and 18-month-old infants and concluded that consistent signs of reliability were mixed, regardless of whether the intervening interval was 1-3 min (31 significant out of 64 correlations) or 1 day (6 significant out of 16 coefficients). Other authors have reported similar small effect size reliabilities in infant behaviors. Notably, Crista et al. (2016) aggregated 0- to 18-day test-retest reliability data from 13 speech perception experiments

conducted in three independent laboratories: Reliability in $N = 409$ 5- to 12-month-olds was “extremely variable” but the weighted mean correlation was negligible ($r = .06$).

Other reports, however, indicate that some infant behaviors are relatively reliable in the short term. Medium to large reliability effect sizes ($r_s = .31$ to $.56$) have been found over sessions held 1 to 3 weeks apart for infant visual habituation/dishabituation measures, such as total looking time, response decrement ratios, trials to criterion, and novelty responses (Bornstein & Benasich, 1986; Fenson, Sapper, & Minner, 1974; Pecheux & Lecuyer, 1983). Fantz (1964) and Bornstein and Benasich (1986) also reported reliability between closely spaced habituation sessions in the patterns of habituation shown by 1- to 6-month-olds and 5-month-olds, respectively. Nozza, Miller, Rossman, and Bond (1991) assessed test-retest reliability in a speech-sound discrimination-in-noise task using a visual reinforcement speech discrimination procedure with an adaptive (up-down) threshold protocol: $N = 16$ 9.5-month-olds provided two thresholds that “in most cases” were within 10 dB of each other (neither the exact inter-assessment interval nor a conventional r were reported). Bornstein, Gaughran, and Seguí (1991) acquired mother and observer ratings of 10 behaviors in $N = 75$ 5-month-olds over two home visits spaced 6 days apart, and both showed large effect size aggregate reliabilities ($r_s = .45$ and $.53$, respectively). Seifer, Sameroff, Barrett, and Krafchuk (1994) later adopted the same approach, having mothers and observers rate behaviors of $N = 50$ 4-month-olds in the home once a week for 8 weeks: Week-to-week correlations were small to medium in effect size ($ICCs = .14$ -. $.36$). Using the Emotional Availability Scales (EAS; Biringen et al., 1998), Bornstein, Gini, Suwalsky, Putnick, and Haynes (2006) found that observed infant responsiveness and involvement had large 1-week reliabilities ($r_s = .50$ and $.48$, respectively) in $N = 52$ 5-month-olds seen in the home. Houston, Horn, Qi, Ting, and Gao (2007, Experiment 3) assessed 1- to 3-day reliability in $N = 10$ 9-month-olds in an audio-video habituation-novelty preference paradigm and found a large novelty-preference reliability ($r = .65$). Maas, Vreeswijk, and van Bakel (2013) reported same-day reliabilities in $N = 292$ 6-month-olds who were videotaped (with their mothers) in three different situations (free play, face-to-face play, and diaper change): Behavioral scales of infant positive mood, negative mood, activity level, sociability, and sustained attention correlated, albeit with great variability, across the three situations ($ICCs = .16$ to $.57$). Munsters, van Ravenswaaij, van den Boomen, and Kemner (2019) reported large short-term reliabilities of overall brain cortical responses (N290, P400, Nc) in $N = 31$ 9- to 10-month-olds who were tested twice within 2 weeks ($r_s = .69$ -. $.77$). van der Velde, Haartsen, and Kemner (2019) estimated the 1-week reliabilities of electroencephalographic connectivity and network characteristics in $N = 60$ 10-month-olds at multiple sites in multiple ways: Overall, reliabilities of global connectivity characteristics were high, more local characteristics showed lower but still acceptable reliabilities, and characteristics calculated with the connectivity matrices of theta and alpha1 frequency bands were most reliable.

In overview, the history of infant homotypic reliability studies published between 1964 and 2019 points to a wide variety of systems having been measured with an equally wide variety of levels of reliability ($r_s = .06$ -. $.77$). It is important to note, first, that correlations of $.06$ and $.77$ mean that the measures involved share 0.36% to 59% of their common variance; obversely, they *do not* share 99.64% to 41% of their common variance. The true reliability of

observed infant behaviors in the extant literature seems decidedly mixed. Second, the literature suggests that different systems enjoy different levels of reliability. This modern picture suggests that little has changed in regard to the reliability of infant behaviors since Shirley's (1933) intensive study of motor development during the first 2 years led her to conclude that "Both constancy and change characterize the personality of the baby" (p. 56).

Reliability in Parenting Practices with Infants.

Given the near universal belief in the importance of parenting, especially in the first years of life, reports of the short-term homotypic reliability of (mothers') parenting practices with infants are (dismayingly) few. Holden and Miller (1999) meta-analyzed repeated behavioral observations of parents (likely mostly mothers) engaging in the same activity at the same location (generally free play in the home or laboratory) over short periods of time (from 3 days to 1 month apart) across 11 studies. They calculated the median reliability correlation at .59. Repeated observations of maternal practices over short periods of time appear to provide a large effect size reliability, and studies of individual parenting practices since have supported this approximate level of reliability. For example, in the EAS study described above Bornstein and colleagues (2006) found that sensitivity and structuring had large 1-week reliabilities ($r_s = .62$ and $.54$, respectively) in $N = 52$ mothers of 5-month infants seen at home, and in the cross-situation study described above Maas et al. (2013) reported that sensitivity to non-distress, stimulation of development, and positive regard in $N = 292$ mothers of 6-month infants correlated across the three same-day situations ($ICCs = .39$ to $.71$).

Again, however, the reliability correlation of .59 means that maternal parenting practices across two measurement points closely spaced in time share only 34.8% of their common variance. And, again, different parenting practices for infants vary in their reliability. For this reason, silently echoing Shirley, Holden and Miller (1999, p. 243) concluded, "the nature of child rearing is simultaneously enduring and different."

This Study in Light of the Extant Literature

The developmental science literature points to homotypic reliability as well as unreliability in individual infant behaviors and maternal parenting practices. However, several noteworthy limitations undermine broader conclusions about these short-term reliabilities, and in this study we attempted to overcome them. First, studies in the extant literature evaluate select single infant behaviors or parent practices, and no one study (to our knowledge) has assessed homotypic reliabilities in a broad range of infant behaviors or maternal parent practices in the same infants or mothers at the same time. In consequence, we do not possess a clear and complete picture of reliabilities of basic infant behaviors or maternal parenting practices all together. For example, as Holden and Miller noted, the median parenting reliability was .59, but across 11 studies the range of reliabilities varied from a low of .35 to a high of .78. Does this range mean that basic individual parenting practices vary in their true reliabilities? Or does it mean that different samples of parents in different studies under different conditions produce varying reliabilities? In the absence of a single omnibus study we do not know. Thus, the generalizability of results in the extant infancy and parenting literatures is in

question. To know which infant behaviors and which maternal parenting practices are reliable, and to what degree vis-à-vis others, an omnibus multivariate approach in the same infants and mothers is needed. We do so here. Moreover, the infant behaviors and maternal parenting practices we studied are universal to infants and parents, respectively, and this study attempted to cover the territory in terms of key developmental and performance competencies that are critical to infant ontogenetic adaptation and the primary parenting tasks of a caregiver of an infant. Second, many psychometric studies of infancy (and of parenting) take place in controlled laboratory situations, especially so assessments of the test-retest reliability of infant capacities (e.g., Crista et al., 2016). Here we studied naturalistic interactions between infants and mothers at home and in doing so attempted to remain faithful to a principle of ecological validity (Bronfenbrenner, 1979; Connors & Glenn, 1996). Furthermore, to localize reliabilities to members of the dyad more precisely we held constant people (only mother present) and infant state (times of the day for observations were selected to provide for favorable assessment conditions, infants were observed to be in states of alertness throughout the course of the observations, and mothers were in the visual presence of their infants). Third, most reliability studies do not take partner characteristics into account. As our design involved mother-infant interactions, to isolate our evaluations of short-term reliabilities of infant behaviors and maternal parenting practices, respectively, we controlled maternal parenting practices in analyzing reliabilities of infant behaviors and we controlled infant behaviors in analyzing reliabilities of maternal parenting practices. Fourth, in contrast to more typical verbal reports or global ratings of infants and mothers, we undertook close quantitative analyses of observed frequencies and durations of individual infant behaviors and maternal parenting practices (called the “gold standard” for assessment; Hawes & Dadds, 2006). Fifth, our assessments of reliability lasted approximately 1 hour on each of two visits and therefore conformed to optimal recording durations (Holden & Miller, 1999, p. 239) and exceeded the durations of many typical assessments of infant behaviors and maternal parenting practices. Sixth, as implied correlation calls for careful interpretation. A “large effect” of, say, .50 (Cohen, 1988) means that 25% of variance is *shared*, but 75% of common variance is *unshared* across two measurements of the same characteristic. We pay close attention to this issue and discuss its significant implications for infant and parenting research.

Although we studied reliability over a short (1-week) period, we hypothesized based on the extant literature, first, that reliabilities of infant behaviors and maternal parenting practices would be small, but also vary by system. We considered it unreasonable to expect that all infant behaviors and all maternal parenting practices that we measured would be equally reliable. Indeed, the amount of variation in reliability expected or considered normal in a characteristic is a function of several factors, including notably the characteristic studied (some characteristics are likely more reliable than others; Maloney & Ward, 1976). Infancy is normally associated with dramatic growth and change in physical motor skills and balance, exploration of the object world, communicative capacities, and socioemotional expressiveness even over short periods, and infants are subject to state fluctuations that might disrupt even short-term measures of reliability (Bornstein et al., 2014). Based on the extant literature, we expected that reliabilities of infant physical development and distress vocalization would be larger – as to the former once infants achieve a motor milestone they

should reproduce it, and as to the latter infant distress vocalization has previously been reported to be moderately consistent (Bornstein et al., 1991). For their part, competing forces shape varying expectations about reliability of maternal parenting. On the one hand, many individual maternal parenting practices have been reported to show temporal consistency. On the other hand, the period following the birth of a first child is unique, and the transition to parenthood entails a host of new experiences and dramatic changes at neurobiological, interpersonal, and ecological levels (Ryan & Padilla, 2019). Based on the extant caregiving literature, we expected to find variability in reliabilities of maternal parenting practices even in the short-term. We hypothesized, second, that infant behaviors will be less reliable than maternal parenting practices. The extant empirical literature indicates that reliability increases with age (Hartmann et al., 2015; Roberts & DelVecchio, 2000), and so reliabilities in adults would be expected to exceed reliabilities in infants. We hypothesized, third, that because parenting plays a formative role in early life (Bornstein, 2019) and mothers organize their infants' behaviors short-term reliability coefficients of infant behaviors might attenuate when maternal parenting practices are taken into account, but that reliabilities of maternal parenting practices would be more refractory to infant behaviors.

Method

Participants

Altogether 51 infant-mother dyads (28 mother-daughter and 23 mother-son dyads) participated in two 1-hr home observations scheduled approximately 1 week apart. Infants were all firstborn, term, free of any known neurological and sensory abnormalities, and healthy at the times of the study. Families were recruited through mass mailings and newspaper advertisements, and private obstetric and pediatric groups, from a large East coast metropolitan area. At the first visit, infants averaged 161.8 days ($SD = 4.5$, $range = 153-173$). The second visit occurred on average 6 days after the first ($M = 6.2$ days, $SD = 2.4$). At birth infants weighed an average of 3.5 kg ($SD = 0.4$). Mothers averaged 29.7 years ($SD = 4.9$) at the time of the study; two mothers had not completed high school, 4 had completed high school only, 10 partial college, 18 college, and 17 had enrolled in or completed university graduate programs. Families ranged from low to high socioeconomic status (SES; Hollingshead, 1975, Four-Factor Index of Social Status, $M = 56.0$, $SD = 10.0$, $range = 16$ to 66). Because these demographic variables were unrelated to any infant or mother measures, they were not considered in the analyses. Infant development and parenting are known to vary with ethnicity (Bornstein & Lansford, 2010; Halgunseth, 2019; McLoyd, Hardaway, & Jocson, 2019; Murry, Hill, Witherspoon, Berkel, & Bartz, 2015; Ng & Wang, 2019). We therefore recruited a sociodemographically heterogeneous, but ethnically homogenous, European American community sample as a first step in understanding reliabilities of infant behaviors and maternal parenting practices. By including only European American infants and mothers, we intentionally avoided an ethnicity confound that might cloud our findings with respect to infant behavior, maternal parenting, and reliability (Bornstein, Jager, & Putnick, 2013; Jager, Putnick, & Bornstein, 2017). In consequence, the generalizability of our findings is specific and clear but limited. Family recruitment and the conduct of this research were approved by the NICHD Institutional

Review Board under Protocol #88-CH-0032 under Title: Specificity of Mother-Infant Interaction.

Home Observation Procedures and Coding

Infants and mothers were visited and audio/videorecorded in their homes for approximately 1 hr of naturalistic ongoing interaction when only infant, mother, and a female researcher were present. Codable time totaled less than 50 min (but more than 45 min) for one dyad, and data from that dyad were prorated. Records were coded using a mutually exclusive and exhaustive continuous and comprehensive coding system yielding unbiased estimates of behavior and practice frequency and duration.

The two reliability visits were scheduled when infants were awake and alert and no other family members were present. Mothers were asked to behave in their usual manner and to disregard the observer's presence insofar as possible. After a standard period of acclimation to the recording equipment and the presence of the observer (McCune-Nicolich & Fenson, 1984; Stevenson, Leavitt, Roach, Chapman, & Miller, 1986), audio/videorecording commenced. The observer refrained from talking to or making eye contact or interacting with or otherwise reacting to the infant or mother during recording.

Infant behaviors and maternal parenting practices and context indicators were categorized into a taxonomy of interactional domains. Five infant domains were identified representing 13 key developmental and performance competencies that are critical to successful ontogenetic adaptation of an infant in the middle of the first year of life: physical development, social interaction, exploration, nondistress vocalization, and distress communication. Appendix 1 lists infant domains, behaviors, interim variables, and final indicator variables. Six maternal domains were identified encompassing the primary parenting tasks required of the mother of a young infant: nurturing, physical and verbal encouragement of motor growth, social exchange, didactic interaction, provision of the material environment, and speech to infant. These domains, parallel to the infant domains, were referenced by 12 behavioral and context indicators. Appendix 2 lists maternal parenting practice domains, practices, interim variables, and final indicator variables. Domain scores were calculated as the mean of the (usually standardized) infant behaviors, maternal parenting practices, and context indicators that related conceptually to the domain but did not need to meet the criteria of an internally consistent scale (see Bradley, 2004; Streiner, 2003). Frequencies and durations of individual behaviors and practices were coded; this microanalytic strategy allowed us to examine infant and mother activities at the level of in-the-moment lived experiences. For behaviors that were continuously coded, *kappa* (κ ; Cohen, 1960, 1968) is reported for intercoder reliabilities based on sec-by-sec agreement; for time-sampled behaviors, the *Intra-Class Correlation* (*ICC*; McGraw & Wong, 1996) is reported.

Infant Behavior Indicators and Domains

The infant *physical* domain score is the mean of two indicators: Infant Balance and Infant Movement, that together assess gross motor development (*ICC* = .96). Underlying the two indicators were four scales of motor ability ordered with respect to their appearance in

ontogeny. The anchor points for each scale, and their developmental equivalents in months, are reported. Sitting was the ability to control the body while in a sitting position ($ICC = .88$). Prelocomotion-upper body was the ability to control and coordinate the upper body while in a prone position ($ICC = .72$). Prelocomotion-lower body was the ability to control and coordinate the lower body while in a prone position ($ICC = .87$). Locomotion was nonaccidental, unassisted movement in any direction, lasting a minimum of 30 continuous sec ($ICC = .95$). For each consecutive 10-min observation interval, the infant was assigned the highest level of each motor ability that was observed. If the infant was never in the physical position necessary to exhibit a given skill during an interval, no rating was made for that skill; in the case of some skills for some infants, no ratings were made in any 10-min interval because the infant was never placed in the required position. The first indicator of the infant physical domain score, Infant Balance, was the highest level of sitting. Consistent with the theoretical understanding that the prelocomotion-upper body, prelocomotion-lower body, and locomotion scales indexed a single dimension of movement, the second indicator of the infant physical domain score, Infant Movement, was defined as the highest level of the prelocomotion-upper body, prelocomotion-lower body, and the locomotion scale scores ($ICC = .88$). The domain score is the mean of the two indicators for which scores were available in at least one 10-min period. As such, it represents the general performance of gross motor functioning in the infant.

Infant *social* is the mean aggregate of the following three indicators ($\kappa = .62$). Look at mother is the mean standard aggregate of the number of times and total duration the infant looked at the mother's face ($\kappa = .67$). Smile is the mean standard aggregate of the number of times and total duration the infant emitted a clear, unambiguous smile ($\kappa = .48$). With respect to the Kappa for infant smile, five points are noteworthy: (a) smiling was an infrequent (and very brief) event; (b) given the considerable disparity in the base rates for the occurrence and nonoccurrence of smiling, the prevalence index (deflating Kappa) was very large = .99; (c) the bias index (inflating Kappa) was very small, maximum = .0007; (d) the proportion of maximum attainable Kappa which was achieved = .95; and (e) Kappa was greater than the less rigorous criterion of .40. Alert expression is the standard score of total length of time the infant's facial expression indicated interest, concentration, staring, or wide-eyed alertness ($\kappa = .72$).

Infant *exploration* is the mean aggregate of the following five indicators ($\kappa = .72$). Look at object is the mean standard aggregate of the number of times and total duration the infant looked at any discrete object or body part other than a face ($\kappa = .71$). Touch object is the mean standard aggregate of the number of times and total duration the infant actively and purposefully handled an object by grasping it and moving it or by directly exploring the object using the palm or fingers of the hand (e.g., patting, rubbing, etc.) ($\kappa = .68$). Mouth object is the mean standard aggregate of the number of times and total duration a discrete object other than a bottle or pacifier came into contact with the infant's mouth ($\kappa = .76$). Extent of exploration is the mean standard aggregate of the variety (the number of different objects), density (the mean number of objects per consecutive 5-min time unit), and consistency (the number of consecutive 5-min time units) of objects the infant mouthed or touched. Efficiency of exploration is the mean standard aggregate of the proportions of

variety, density, and consistency of objects the infant mouthed or touched (number of objects explored, $ICC = .93$).

Infant *vocalization* is a single indicator ($\kappa = .70$). Nondistress vocalization is the mean standard aggregate of the number of times and total duration the infant emitted a positively or neutrally toned vocalization.

Infant *distress communication* is the mean standard aggregate of two indicators ($\kappa = .68$). Negative facial expression is the mean standard aggregate of the number of times and total duration the infant displayed a distressed, angry, or frowning countenance ($\kappa = .67$). Distress vocalization is the mean standard aggregate of the number of times and total duration the infant emitted vocalizations that indicated protest, anger, complaint, or upset ($\kappa = .70$).

Maternal Parenting Practice Indicators and Domains

Mother *nurture* is the mean standard aggregate of the following three indicators ($\kappa = .91$). Feed/burp/wipe is the sum of the durations of two behaviors: the total length of the time the mother fed her infant and burped or wiped her infant ($\kappa = .92$). Bath/diaper/dress/groom/other health needs is the sum of the durations of five behaviors: the total length of time the mother bathed the infant, checked or changed the infant's diaper, dressed the infant, groomed the infant, and attended to the infant's health needs ($\kappa = .88$). Hold is the total length of time the mother supported some or all of her infant's weight with her body ($\kappa = .93$).

Mother *physical* is the mean of the following two indicators ($ICC = .61$). Encourage balance is the mean proportion of consecutive 10-min intervals in which the mother physically or verbally encouraged her infant to sit or stand ($ICC = .51$). Encourage movement is the mean proportion of consecutive 10-min intervals in which the mother physically or verbally encouraged her infant to roll, crawl, or step ($ICC = .83$).

Mother *social* is the mean aggregate of the following three indicators ($\kappa = .70$). Encourage attention to mother is the mean standard aggregate of the number of times and total duration the mother attempted to draw her infant into face-to-face interaction with herself ($\kappa = .71$). Social play is the mean standard aggregate of the number of times and total duration the mother verbally or physically amused the infant, for example, to elicit a smile, positive vocalization, laughter, or motoric excitement ($\kappa = .77$). Express affection physically or verbally is the mean standard aggregate of the number of times and total duration the mother showed affection or positive evaluation to her infant ($\kappa = .66$).

Mother *didactic* is a single indicator ($\kappa = .73$). Encourage attention to objects is the mean standard aggregate of the number of times and total duration the mother physically moved her infant or an object so that her infant could see or touch it or verbally referred to an object-related event or activity.

Mother *material* is the mean aggregate of the following two indicators ($ICC = .91$). Quantity of objects is the mean standard aggregate of the variety (the number of different objects within infant reach), density (the mean number of objects within infant reach per consecutive 5-min time unit), and consistency (the number of consecutive 5-min time units

in which any object was within infant reach) of toys, books, and household objects that were within the infant's reach ($ICC = .94$). Quality (responsiveness) of objects is the mean standard aggregate of the responsiveness of the objects, number of highly responsive objects, and proportion of highly responsive objects within reach of the infant ($ICC = .87$).

Mother *language* is a single indicator ($\kappa = .69$) of speech to infant. The mean standard aggregate of the number of times and total duration the mother used adult-directed speech (i.e., normal intonation patterns) and infant-directed speech (i.e., speech marked by short sentences, repetition, and high and more variable intonation).

Analytic Plan

Prior to all analyses, univariate distributions for all variables were checked for normality and outliers (Tabachnick & Fidell, 2012), and pairs of repeated measures were examined for influential bivariate outliers by scatter plot inspection and numeric statistics (the studentized deleted residual and Cook's D). Transformed variables were used in analyses; for clarity, untransformed data are presented in reports of descriptive statistics.

In addition to zero-order correlations, we evaluated controlled reliability by removing shared variance of corresponding mother parenting practices from infant behaviors, and corresponding infant behaviors from mother parenting practices. Controlled correlations determine whether reliability in one member of the dyad is a function of behaviors/practices in the other member of the dyad. For example, if reliability does not attenuate for an infant behavior when controlling the corresponding maternal practice, it indicates that the infant behavior was reliable and mothers were not driving reliability of the infant behavior. Corresponding partner covariates were those with hypothesized conceptual relations: (a) infant physical and mother physical, (b) infant social and mother social, (c) infant social and mother language, (d) infant exploration and mother didactic, (e) infant exploration and mother material, and (f) infant nondistress vocalization and mother language. To qualify as a covariate, the corresponding infant behaviors and maternal parenting practices had to correlate significantly ($p < .05$) and meaningfully (share at least 5% of the variance) with each other. The 5% rule was adopted because we were only interested in controlling for variables that were practically important as well as conceptually compelling. When a significant and meaningful correlation was found, the residual from a linear regression of the infant behavior or maternal parenting practice on the corresponding practice or behavior, respectively, was computed and used in the controlled correlation analysis.

Post-hoc power analysis (Faul, Erdfelder, Lang & Buchner, 2007) indicated a 99.22% chance of detecting a large effect size and a 71.56% chance of detecting a medium effect size for a one-tailed test significant at the .05 level.

Results

Descriptive statistics for coded behaviors/practices, domain indicators, and domains appear in Table 1. If an observed behavior/practice was standardized to be included in the domain or indicator, the unstandardized variable is represented in Table 1. Table 2 presents short-term reliabilities for infant and mother domain scores and indicators.

Overall, infant behavior domains showed medium to large effect size reliabilities across the 6 days that separated the two home visits. Three out of 5 infant domains were reliable at a medium effect size, mean r of all domains = .30, $p = .02$ (one-tailed test). Only the infant Social and Vocalization domains failed to show short-term reliability. Reliabilities of the infant domain indicators varied widely from $r = -.02$ to .74. The average correlation of the infant indicators was $r = .38$.

Overall, maternal parenting practice domains showed medium to large effect size reliabilities across the 6 days that separated the two home visits. Five of 6 maternal domains were reliable at a large effect size, mean r of all domains = .48, $p < .001$ (one-tailed test). Only the mother Material domain failed to show short-term reliability. Reliability of the mother domain indicators varied widely from $r = -.08$ to .63. The average correlation of the mother indicators was $r = .42$.

We underscore, however, that on average infant domains shared only 9% common variance and maternal domains only 23% common variance over this short time. Moreover, when corresponding maternal practices were controlled, only the infant Physical domain remained statistically reliable across time. Besides the mother Material domain, which was not reliable at the zero-order level, maternal parenting practices remained reliable after removing shared variance with corresponding infant behaviors.

Discussion

Developmental science is broadly interested in individual variation and in ascertaining to what degree children and parents are consistent (or not) in their behaviors and practices, respectively. Reliability describes the situation where individuals in a cohort are consistent relative to one another over short time intervals. As reviewed, there are fundamental practical, theoretical, substantive, clinical, and methodological reasons to evaluate reliability of infant behaviors and maternal parenting practices: Each is descriptive, explanatory, and predictive in its own way. In this prospective longitudinal investigation, we examined short-term reliabilities of a wide swath of infants' behaviors and of mothers' parenting practices. The results reveal in the first year of life (1) small to large effect size short-term reliabilities in infant behaviors, (2) small to large effect size short-term reliabilities in maternal parenting practices, (3) controlling maternal parenting practices from infant behaviors exerts attenuating effects on reliability in infants, (4) controlling infant behaviors from maternal parenting practices hardly affects reliabilities of parenting practices, however (5) even statistically significant reliability correlations in infants and mothers leave considerable amounts of autoregressive shared variance unaccounted for.

Reliability of Infant Behaviors

This study considered short-term reliability of more than a dozen infant behaviors across five domains of physical, social, exploration, nondistress vocalization, and distress communication and found that over 1 week three of five domains were reliable at medium to large effect sizes. The mean overall short-term reliability correlation achieved a medium effect size of .30, but a mean overall short-term reliability correlation of .30 means that 90% of common variance in infants' behaviors was unshared even over so short a period of time

as 6 days. Notably, too, only one domain remained reliable in infants when mothers' parenting was taken into account. The diminished reliability of infant behaviors when controlling for maternal practices suggests that mothers play a formative role in supporting the consistency of infant behaviors in the middle of the first year of life. Infancy is often thought of as a highly variable phase of the life course with behaviors fluctuating from moment to moment; our reliability and controlled data reinforce that perspective and indicate that some common consistency in infants may be accounted for by their mothers.

Reliability of Maternal Parenting Practices

Maternal parenting was on average reliable at a large effect size (.48), and only the material domain in parenting was not reliable. (This exception is striking because one would not think the infant's environment changes over so short a period of time as 1 week. Based on the indicator *rs*, it appears that quality of objects changes more than quantity; it may be, therefore, that mothers change the toys infants interact with across observations.) These reliability findings in maternal practices accord with Holden and Miller's (1999) 11-study meta-analysis. Again, however, the average 83% unshared variance in maternal parenting practices is large. As predicted and as has been found previously in studies of long-term stability (e.g., Kochanska & Aksan, 2004; NICHD Early Child Care Research Network, 1999, 2003; Weinfield et al., 2002), the overall average maternal parenting practices reliability (.48) was larger than the overall average infant behaviors reliability (.30).

Causes and Consequences of Reliability

Obtained reliability scores may be decomposed into two generic components: true reliability and error. The true score portion of the obtained score ordinarily is the part that remains constant across time (Hartmann et al., 2015). True scores in development (reliability included) are governed by genetic and biological factors inextricably intertwined with influences of environment and experience. Thus, genetic and biological characteristics of infants and mothers could underpin consistencies in their behaviors and practices (Pérusse, Neale, Heath, & Eaves, 1994; Saudino, 2012; Broderick & Neiderhiser, 2019), just as stable environmental and experiential characteristics of infants and mothers could promote consistencies in their behaviors and practices (Belsky & Isabella, 1988; Bradley, 2019). These twin life forces are indissociable, and therefore true reliability of any characteristic is likely attributable to their transactions (e.g., Bornstein, 2019; Sameroff, 2009).

The discrepancy between an obtained reliability coefficient and perfect reliability (1.00) is an index of the relative amount of measurement error. The error component of reliability is the portion of the score that changes across time and results in unreliable performance. Error too has many possible sources and explanations. Unreliable performance might be produced by chance events or the assessment setting (uncontrolled aspects of the conditions under which reliability is assessed, changing availability of people or objects in the environment), temporary states of the participant (fluctuations in state, systematic oscillations), idiosyncratic aspects of the measurement instrument or procedures (sampling the domain of content, inconsistent observer behavior or scorer error, recording or coding), or the characteristic itself (normal fluctuations in true scores, real change). The nature of psychological measurement precludes the possibility of perfect reliability, of course, and an

amount of variation in reliability is expected and considered normal. In measurement, the goal is to keep unreliability to a minimum. Our recording was standardized and our coding ensured psychometrically adequate measurement. The question of how high reliability should be is difficult as there is no hard and fast criterion (Maloney & Ward, 1976, p. 63). That said, conventional psychometric theory suggests rules of thumb, and a reliability coefficient of .70 is generally required for a test to be considered an acceptable measure of a trait in individual-difference research (Cronbach, 1951; Nunnally, 2017; Wiggins, 1973). Only maternal Language achieved this criterion.

Strengths and Limitations

First, our participants were European American 5-month infants and their primiparous mothers. This sampling is by no means invalid—the population to whom the findings generalize is clear — but may have implications for any broader generalizability of the findings. Different patterns of reliability could emerge in multiparous or single mothers, at-risk samples, in other ethnic or cultural groups, or for that matter in fathers, alloparents, or other caregivers. For example, parenting is moderated by mothers' adolescent versus adult status (Lounds, Borkowski, Whitman, Maxwell, & Weed, 2005), their SES (Jenkins et al., 2003), as well as their culture (Bornstein, Tamis-LeMonda, Tal, Ludemann, Toda, Rahn, Pêcheux, Azuma, & Vardi, 1992). A related implication of our sampling is that the reliabilities we found may underestimate “true” reliabilities of these behaviors and practices because more homogeneous samples (like ours) may restrict between-family variance which attenuates reliability and because our estimates of reliability derive from associations between observations that are not themselves perfectly reliable. We concentrated on selected infant behaviors and maternal parenting practices occurring in open interactions based on fixed procedures and durations; we also examined infant behaviors and maternal parenting practices in a relatively narrow, if still developmentally significant, time window at 5 months. Whether other infant behaviors or maternal parenting practices (e.g., feeding, punishment) would yield different degrees of reliability, do so under other testing parameters (e.g., shorter and longer assessments are known to yield smaller effect size reliabilities, but shorter interassessment intervals are known to yield larger effect size reliabilities than longer ones), or do so at earlier or later developmental periods is open to question (e.g., reliabilities with older children might be greater than with infants). That said, the behaviors and practices we studied are universal to infants and parents, respectively. When during the day infants and mothers were observed also provided for favorable assessment conditions: Infants were observed to be in states of alertness throughout the course of the observations, and mothers were in the visual presence of their infants. Furthermore, we assessed behaviors and practices; verbal reports are known to enjoy higher reliabilities than observations (Holden & Miller, 1999). We also investigated one model of reliability, *homotypic*; that is, maintenance of order amongst the same individuals on the same characteristic over time ($A \rightarrow A$). A complementary model describes *heterotypic reliability*, the maintenance of order amongst the same individuals on different (if related) manifest characteristics through time ($A \rightarrow A'$). Likely, homotypic reliabilities enjoy larger effect sizes than heterotypic reliabilities.

Conclusions

The present study is concerned with understanding the nature and scope of short-term reliability in infancy. Persistent and systematic child-rearing practices are often credited with affording experiences that influence the course and outcome of infant development (e.g., Bornstein, 2019; Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2001; Maccoby, 2000; Vandell, 2000). As Maccoby (1984, p. 326) observed, “the family system, like any system, has self-stabilizing properties ... families tend to stabilize around habitual patterns of interaction; thus, there is continuity over time in... familial forces...” Submitted to empirical test, however, we found a range of reliabilities with consistently small amounts of shared variance even between short intervals. Infants who are consistent in their behaviors or who engender or experience consistent parenting practices likely follow consistent developmental paths, whereas infants who are inconsistent in their behaviors or who engender or experience inconsistent parenting practices likely follow divergent ontogenetic paths. A principal charge of developmental science is to document how reliabilities might condition these ontogenetic trajectories. In this regard, we note that “unreliability” has *paradoxically* distinct advantages, especially in infancy: It could be that unreliability in infants and mothers, which sounds undesirable and deleterious to development, in actuality reflects flexibility and openness to adapting to new and novel environmental demands and experiences. On this argument, developmental science needs to move beyond exclusively focusing on reliability, as reliability, variability in reliability, and unreliability of infant behaviors and maternal parenting practices all inform our understanding infancy and parenting. We think appreciating the origins of each and their implications for child development will pay dividends.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the NIH/NICHD, USA, and an International Research Fellowship at the Institute for Fiscal Studies (IFS), London, UK, funded by the European Research Council (ERC) under the Horizon 2020 research and innovation programme (grant agreement No 695300-HKADeC-ERC-2015-AdG).

Appendix

Appendix 1

Infant Domains, Behaviors, Interim Variables, and Final Indicator Variables

Domain	Behavior: Definition	Interim Variable(s)	Final Indicator Variable
Physical ^a	<i>Sit</i> : Infant may be placed in initial sitting position, but ability to maintain and control sitting balance is evaluated, with control lasting a minimum of 30 continuous seconds. The rating scale contained 8 levels plus “Not coded” (no opportunity to observe) from Level 1: Sits with back rounded & head unsteady (bobs, leans to side, falls forward) when fully supported in inclined sitting position - adults lap, infant seat, and so forth. to Level 8: Rotates from prone position to a balanced sitting position with weight on buttocks and without assistance.	The score for sit was converted to its equivalent developmental level expressed in months.	Balance is the highest developmental level score for sit observed in five consecutive 10-minute time units.

Domain	Behavior: Definition	Interim Variable(s)	Final Indicator Variable
	<i>Prelocomotion, Upper Body:</i> The prone infant lifts the head and shoulders and/or extends the arms, lasting a minimum of 30 continuous seconds unless otherwise noted. The rating scale contained 5 levels plus "Not coded" (No opportunity to observe) from Level 1: The prone infant lifts head and shoulders for 5+ sec.; arms not used as primary support. to Level 5: The prone infant, up on extended arms and able to reach with one arm, shifting weight and remaining balanced. Well-coordinated movements.	Upper body prelocomotion, lower body prelocomotion, and locomotion scores were initially converted to their equivalent developmental levels expressed in months. Then a movement score expressed in months was computed for each of the five 10-minute observational periods as the highest of the developmental level scores for upper body prelocomotion, lower body prelocomotion, and locomotion.	Movement is the highest movement score expressed in months observed in five consecutive 10-minute time units.
	<i>Prelocomotion, Lower Body:</i> The prone infant extends/lifts hips, bends knees, and supports weight on knees and lower legs, lasting a minimum of 30 continuous seconds. The rating scale contained 4 levels plus "Not coded" (No opportunity to observe) from Level 1: The prone infant extends legs with hips resting on supporting surface. to Level 4: The prone infant supports weight on knees and lower legs and off thighs. (Full crawl position, with or without movement.)		
	<i>Locomotion:</i> The infant displays "deliberate" or nonaccidental, unassisted movement in any direction lasting a minimum of 30 continuous seconds, with or without a noted "goal." The rating scale contained 11 levels plus "Not coded" (No opportunity to observe) from Level 1: The infant lifts his/her legs when in a supine position, with or without attempts to grasp the feet or legs; the arms and legs are active. to Level 11: The infant actively creeps across the room.		
Social	<i>Look at mother:</i> The infant looks at the mother's face or head. Focused fixation must be evident. An active behavior component often accompanies clear and focused fixation (e.g., brightening of the face, widening of the eyes, stilling, increased motor excitement, positive vocalizations, or reaching). A change in fixation is coded after the infant has looked away from target for 1 second.	Frequency of looking at mother Duration of looking at mother	Mean standard score of frequency and duration of looking at mother.
	<i>Smile:</i> The infant emits a clear, unambiguous smile. The corners of the baby's mouth are extended outward and upward; the eyes 'brighten' and are focused; and the eyebrows are relaxed or raised.	Frequency of smiling Duration of smiling	Mean standard score of frequency and duration of smiling.
	<i>Alert expression:</i> The infant's face lacks clear indications of positive or negative affect. This category includes such facial poses as expressions of interest, concentration or seriousness, questioning looks, and wide-eyed alertness.	Duration of alert expression	Mean standard score of duration of alert expression.
Exploration	<i>Look at object:</i> The infant looks at any discrete object or body part other than a face that is within a radius of 12 feet. Focused fixation must be evident. An active behavior component often accompanies clear and focused fixation (e.g., brightening of the face, widening of the eyes, stilling, increased motor excitement, positive vocalizations, or reaching). A change in	Frequency of looking at object Duration of looking at object	Mean standard score of frequency and duration of looking at object.

Domain	Behavior: Definition	Interim Variable(s)	Final Indicator Variable
	fixation is coded after the infant has looked away from target for 1 second.		
	<i>Touch object</i> : The infant actively and purposefully handles an object by grasping and moving it (e.g., lifting, waving, banging, dropping, rotating, and turning) or by directly exploring the object using the palm or fingers of the hand (e.g., patting, rubbing, squeezing, and fingering). A change in behavior is coded when the new behavior has lasted for 1 second.	Frequency of touching object Duration of touching object	Mean standard score of frequency and duration of touching object.
	<i>Mouth object</i> : A discrete object other than a pacifier or bottle is in contact with the infant's mouth. A change in behavior is coded when the new behavior has lasted for 1 second.	Frequency of mouthing object Duration of mouthing object	Mean standard score of frequency and duration of mouthing object.
	<i>Extent of exploration</i> : The amount of touching or mouthing an infant does of toys, books, or household objects that are within reach.	<i>Variety of objects explored</i> : The number of different objects explored during the total observation. <i>Density of objects explored</i> : The mean number of objects explored per 5-minute time unit. <i>Consistency of objects explored</i> : The number of consecutive 5-minute time units in which any object was explored.	Mean standard score of variety, density, and consistency of objects explored.
	<i>Efficiency of exploration</i> : The proportion of toys, books, or household objects which are within reach that an infant touches or mouths.	<i>Proportion variety</i> : The proportion of available objects explored during the total observation. <i>Proportion density</i> : The mean proportion of available objects explored per 5-minute time unit. <i>Proportion consistency</i> : The proportion of consecutive 5-minute time units containing available objects in which exploration occurred.	Mean standard score of proportion variety, proportion density, and proportion consistency.
Vocalization	<i>Nondistress vocalization</i> : Any positively or neutrally toned infant vocalization that is clearly audible. Included are babbling, cooing, laughing, vocal play, shrieking, and sighs or grunts not indicative of distress. Vocalizations of any duration are coded; brief pauses in vocalization of less than 1 s are not recorded.	Frequency of nondistress vocalization Duration of nondistress vocalization	Mean standard score of frequency and duration of nondistress vocalization.
Distress communication	<i>Negative facial expression</i> : The infant displays a distressed, angry, disgusted, or frowning expression, characterized by at least two of the following: mouth opened and lips stretched horizontally, mouth closed and lips pressed together, an inward and downward pressing of the bridge of the nose, and one or more horizontal furrows across the forehead, especially just above the eyebrows.	Frequency of negative facial expression Duration of negative facial expression	Mean standard score of frequency and duration of negative expression.
	<i>Distress vocalization</i> : Vocalizations produced by the infant that indicate protest,	Frequency of distress vocalization	Mean standard score of frequency and

Domain	Behavior: Definition	Interim Variable(s)	Final Indicator Variable
	complaint, anger, or upset, as indicated by vocal quality, facial expression, or other negative behaviors (e.g., intense squirming, back arching). Vocalizations of any duration are coded; brief pauses in vocalization of less than 1 s are not recorded.	Duration of distress vocalization	duration of distress vocalization.

Note.

^aEach of the scales of infant motor skills consisted of a hierarchy of operationalized behavioral abilities, ordered from least mature to most mature and converted to the equivalent developmental level expressed in months (Bly, 1981). For a time unit in which the infant was never in the physical position necessary to exhibit a given skill, the skill was coded "No opportunity to observe," and no rating was assigned.

Appendix 2

Mother Domains, Parenting Practices, Interim Variables, and Final Indicator Variables

Domain	Parenting Practice: Definition	Interim Variable(s)	Final Indicator Variable
Nurture	<i>Feed</i> : The mother attempts to give the infant liquid or solid foods by cup, bottle, breast, or spoon.	Duration of feed	Sum of durations of feed and burp/wipe.
	<i>Burp/Wipe face or hands</i> : The mother attempts to burp the infant in connection with a feeding, or the mother wipes the infant's face, hands, or clothing at any time.	Duration of burp/wipe	
	<i>Bathe</i> : The mother washes and dries the infant's body and/or hair.	Duration of bathe	Sum of durations of bathe, check/change diaper, dress, groom, and meet other health needs.
	<i>Check/Change diaper</i> : The mother checks to see if the infant needs a diaper change or changes the diaper.	Duration of check/change diaper	
	<i>Dress</i> : The mother removes or puts an article of clothing on the infant	Duration of dress	
	<i>Groom</i> : The mother engages in behavior designed to enhance the infant's appearance (e.g., combs hair).	Duration of groom	
	<i>Meet other health needs</i> : The mother attends to other health needs of the infant (e.g., wipes or suctions the infant's nose; gives medicine from a dropper or medicine spoon).	Duration of meet other health needs	
	<i>Hold</i> : The mother supports some or all of the infant's weight with her body.		Duration of hold.
Physical	<i>Physically encourage to sit</i> : The mother places the infant in a sitting position in which the infant's back is not leaning against a firm surface.	Proportion of consecutive 10-minute time units physical encouragement to sit was observed.	Mean proportion of consecutive 10-minute time units in which physical encouragement to sit or stand was observed.
	<i>Physically encourage to stand</i> : The mother places or holds the infant in a standing position so that there is some weight supported by the infant's straightened legs.	Proportion of consecutive 10-minute time units physical encouragement to stand was observed.	
	<i>Physically encourage to roll</i> : The mother physically assists the infant to roll over.	Proportion of consecutive 10-minute time units physical encouragement to roll was observed.	Mean proportion of consecutive 10-minute time units in which physical encouragement to roll, crawl, or step was observed.
	<i>Physically encourage to crawl</i> : The mother physically assists the infant to move forward (on the belly or on hands and knees) by moving the infant's arms and/or legs or by pushing rump or feet from behind.	Proportion of consecutive 10-minute time units physical encouragement to crawl was observed.	

Domain	Parenting Practice: Definition	Interim Variable(s)	Final Indicator Variable
	<i>Physically encourage to step:</i> The mother holds the infant in a standing position and then moves the infant's body to simulate stepping movements.	Proportion of consecutive 10-minute time units physical encouragement to step was observed.	
Social	<i>Encourage attention to mother:</i> The mother attempts to draw the infant into face-to-face social interaction with herself. Physical attempts include intentionally moving her face toward the infant or moving the infant toward her face. Verbal attempts include making very specific comments about herself that are clearly designed to capture the infant's interest. Pauses of 2 seconds or longer are coded as terminations of an ongoing behavior.	Frequency of encouraging attention to mother Duration of encouraging attention to mother	Mean standard score of frequency and duration of encourage attention to mother.
	<i>Social play:</i> The mother directs verbal or physical behavior to the infant, the purpose of which appears to be to amuse the infant (i.e., to elicit smiles, positive vocalizations, laughter, or motoric excitement in the context of a primarily social dyadic interaction). Coding is discontinued when the mother has not interacted for 3 seconds and when she is no longer oriented to the infant and poised to continue the exchange; pauses of any duration when the mother clearly remains poised to continue are coded as part of the play sequence. The types of exchanges coded as social play are: (a) physical contact with a fun-like quality (e.g., tickling); (b) introducing the element of surprise, suspense, or quick release of stimuli (e.g., peek-a-boo); (c) singing to the infant; and (d) playing a game that involves physical manipulation of the infant's body (e.g., patty cake).	Frequency of social play Duration of social play	Mean standard score of frequency and duration of social play.
	<i>Express affection.</i> The mother expresses affection or positive evaluation to the infant either physically (e.g., kissing, patting, stroking, or caressing) or verbally (using explicit phrases denoting praise or endearment).	Frequency of positive affect or evaluation Duration of positive affect or evaluation	Mean standard score of frequency and duration of positive affect or evaluation.
Didactic	<i>Encourage attention to object:</i> The mother physically moves the infant or an object so that the infant can see or touch it, or the mother verbally refers to an object or an object-related event or activity that is no more than 12 feet from the infant. Pauses of 2 seconds or longer are coded as terminations of an ongoing behavior.	Frequency of encouraging attention to object Duration of encouraging attention to object	Mean standard score of frequency and duration of encouraging attention to object.
Material	<i>Quantity of objects provided infant:</i> The number of toys, books, and household objects that are within the infant's reach.	<i>Variety of objects provided:</i> The number of different objects that are within infant reach during the total observation. <i>Density of objects provided:</i> The mean number of objects within infant reach per consecutive 5-minute time unit. <i>Consistency of objects provided:</i> The number of consecutive 5-minute time units in which any object was within infant reach.	Mean standard score of variety, density, and consistency of objects provided.
	<i>Quality (responsiveness) of objects provided infant:</i> Ratings are made of all toys, books,	<i>Responsiveness:</i> Mean of sums of ratings for moving	Mean standard score of responsiveness of objects,

Domain	Parenting Practice: Definition	Interim Variable(s)	Final Indicator Variable
	and household objects within reach of infant on four dimensions: moving parts, change in shape or contour, noise production, and reflected image.	parts, change in shape or contour, noise production, and reflected image for each object within infant reach. <i>Number of highly responsive objects:</i> Number of objects within infant reach that has a sum of responsiveness ratings 12 (on a scale of 4 to 16). <i>Proportion of highly responsive objects:</i> Proportion of objects within infant reach with a sum of responsiveness ratings 12.	number of highly responsive objects, and proportion of highly responsive objects.
Language	<i>Adult-directed speech:</i> Words and speech-like sounds directed by the mother to the infant that are characterized by normal intonation patterns. Included are syllable sounds, parts of words, single words, conversations, and singing. Changes and pauses in vocalization lasting less than 1 second are not recorded.	Frequency of the sum of adult-directed and child-directed speech	Mean standard score of frequency and duration of the sums of adult-directed and child-directed speech.
	<i>Child-directed speech:</i> The special speech register used by mother when talking to her infant, including short sentences, greater repetition and questioning, and higher and more variable intonation than that of speech addressed to adults. Changes and pauses in vocalization lasting less than 1 second are not recorded.	Duration of the sum of adult-directed and child-directed speech	

References

- Alder AG, & Scher SJ (1994). Using growth curve analysis to assess personality change and stability in adulthood In Heatherton TF & Weinberger JL (Eds.), *Can personality change?* (pp. 149–173). Washington, DC: American Psychological Association Doi: 10.1037/10143-007
- Anastasi A (1968). *Psychological testing* (3e). New York: McMillan.
- Baltes PB, & Nesselroade JR (1979). History and rationale of longitudinal research In Nesselroade JR & Baltes PB (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic Press.
- Belsky J, & Isabella R (1988). Maternal, infant, and social-contextual determinants of attachment security In Belsky J & Nezworski T (Eds.), *Clinical implications of attachment*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Biringen Z, Robinson J, & Emde R (1998). *Emotional Availability Scales*, 3rd Edition. Unpublished manual for the EAS-training. Taken from www.emotionalavailability.com
- Bornstein MH (2019). Parenting infants In Bornstein MH (Ed.), *Handbook of parenting*. Vol. 1. Children and parenting (3e, pp. 3–66). New York: Routledge Doi: 10.4324/9780429440847-1
- Bornstein MH, & Benasich AA (1986). Infant habituation: assessments of individual differences and short-term reliability at five months. *Child Development*, 57, 87–99. Doi: 10.2307/1130640 [PubMed: 3948596]
- Bornstein MH, & Lansford JE (2010). Parenting In Bornstein MH (Ed.), *The Handbook of Cultural Developmental Science*. Part 1. Domains of Development across Cultures (pp. 259–277). New York, NY: Psychology Press.
- Bornstein MH, Arterberry ME, & Lamb ME (2014). *Development in infancy: A contemporary introduction* (5e). New York, NY: Psychology Press.
- Bornstein MH, Gaughran JM, & Seguí I (1991). Multimethod assessment of infant temperament: Mother questionnaire and mother and observer reports evaluated and compared at five months

- using the Infant Temperament Measure. *International Journal of Behavioral Development*, 14, 131–151. Doi: 10.1177/016502549101400202
- Bornstein MH, Gini M Suwalsky JTD, Putnick DL, & Haynes OM Emotional availability in mother-child dyads: Short-term stability and continuity from variable-centered and person-centered perspectives. *Merrill-Palmer Quarterly*, 2006, 52, 547–571. Doi: 10.1353/mpq.2006.0024
- Bornstein MH, Jager J, & Putnick DL (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review*, 33, 357–370. doi: 10.1016/j.dr.2013.08.003 [PubMed: 25580049]
- Bornstein MH, Putnick DL, & Esposito G (2017). Continuity and stability in development. *Child Development Perspectives*, 11, 113–119. Doi: 10.1111/cdep.12221 [PubMed: 29503666]
- Bornstein MH, Tamis-LeMonda CS, Tal J, Ludemann P, Toda S, Rahn CW, Pêcheux M-G, Azuma H, & Vardi D (1992). Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development*, 63, 808–821. Doi: 10.1111/j.1467-8624.1992.tb01663.x [PubMed: 1505242]
- Bradley RH (2004). Chaos, culture, and covariance structures: A dynamic systems view of children's experiences at home. *Parenting: Science and Practice*, 4, 243–257. Doi: 10.1207/s15327922par0402&3_7
- Bradley RH (2019). Environment and parenting In Bornstein MH (Ed.), *Handbook of parenting Vol. 2: Biology and ecology of parenting* (3rd edition, pp. 474–518). New York, NY: Routledge Doi: 10.4324/9780429401459-15
- Broderick AV, & Neiderhiser JM (2019). Genetics and parenting In Bornstein MH (Ed.), *Handbook of parenting Vol. 2: Biology and ecology of parenting* (3rd edition). New York, NY: Routledge Doi: 10.4324/9780429401459-4
- Bronfenbrenner U (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Cairns RB (1979). *Social development: The origins and plasticity of interchanges*. San Francisco: Freeman.
- Cannon WB (1932). *The wisdom of the body*. New York, NY: W.W. Norton & Company, Inc.
- Coates B, Anderson EP, & Hartup WW (1972). The stability of attachment behaviors in the human infant. *Developmental Psychology*, 6, 231–237. Doi: 10.1037/h0032088
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. doi: 10.1177/001316446002000104
- Cohen J (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70, 213–220. doi: 10.1037/h0026256 [PubMed: 19673146]
- Cohen J (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Collins WA, Maccoby EE, Steinberg L, Hetherington EM, & Bornstein MH (2000). Contemporary research on parenting: The case for nature *and* nurture. *American Psychologist*, 55, 218–232. Doi: 10.1037/0003-066X.55.2.218 [PubMed: 10717969]
- Connors E, & Glenn SM (1996). Methodological considerations in observing mother-infant interactions in natural settings Haworth J (Ed.), *Psychological research: Innovative methods and strategies* (pp. 139–152). London: Routledge.
- Cristia A, Seidl A, Singh L, & Houston D (2016). Test–Retest Reliability in Infant Speech Perception Tasks. *Infancy*, 21(5), 648–667. DOI: 10.1111/inf.12127
- Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. Doi: 10.1007/BF02310555
- DeVellis RF (2016). *Scale development: Theory and applications*. Thousand Oaks, CA: SAGE Publications, Inc.
- Epstein S (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology*, 37, 1097–1282. Doi: 10.1037/0022-3514.37.7.1097
- Erikson EH (1963). *Childhood and society*. New York: Norton.
- Fantz RL (1964). Visual experience in infants: decreased attention familiar patterns relative to novel ones. *Science*, 146(3644), 668–670. Doi: 10.1126/science.146.3644.668 [PubMed: 14191712]

- Faul F, Erdfelder E, Lang AG, & Buchner A (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. Doi: 10.3758/BF03193146 [PubMed: 17695343]
- Fenson L, Sapper V, & Minner DG (1974). Attention and manipulative play in the one-year-old child. *Child Development*, 45(3), 757–764. Doi: 10.2307/1127842 [PubMed: 4143830]
- Fleeson W (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13, 83–87. Doi: 10.1111/j.0963-7214.2004.00280.x
- Halgunseith LC (2019). Latino and Latin American Parenting In Bornstein MH (Eds.), *Handbook of Parenting*. Vol 4 Social Conditions and Applied Parenting (3e). New York: Routledge Doi: 10.4324/9780429398995-2
- Hartmann DP, Abbott C, & Pelzel K (2015). Design, measurement, and analysis in developmental research In Bornstein MH & Lamb ME (Eds.), *Developmental science: An advanced textbook* (7th ed., pp. 113–214). New York: NY: Taylor & Francis.
- Hawes D, & Dadds M (2006). Assessing parenting practices through parent-report and direct observation during parent-training. *Journal of Child and Family Studies*, 15, 554–567. Doi: 10.1007/s10826-006-9029-x.
- Holden GW, & Miller PC (1999). Enduring and different: A meta-analysis of the similarity in parents' child rearing. *Psychological Bulletin*, 125, 223–254. doi: 10.1037/0033-2909.125.2.223 [PubMed: 10087937]
- Hollingshead AB (1975). The four-factor index of social status. Unpublished manuscript. Yale University.
- Houston DM, Horn DL, Qi R, Ting JY, & Gao S (2007) Assessing Speech Discrimination in Individual Infants *Infancy*, 12(2), 119–145. Doi: 10.1111/j.1532-7078.2007.tb00237.x
- Jager J, Putnick DL, & Bornstein MH (2017). II. More than just convenient: the scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development*, 82(2), 13–30. doi: 10.1111/mono.12296 [PubMed: 28475254]
- Jenkins JM, Rasbash J, & O'Connor TG (2003). The role of the shared family context in differential parenting. *Developmental Psychology*, 39, 99–113. doi: 10.1037/0012-1649.39.1.99 [PubMed: 12518812]
- Kochanska G, & Aksan N (2004). Development of mutual responsiveness between parents and their young children. *Child Development*, 75(6), 1657–1676. 10.1111/j.1467-8624.2004.00808.x [PubMed: 15566371]
- Landis JR, & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. 10.2307/2529310 [PubMed: 843571]
- Lounds JJ, Borkowski JG, Whitman TL, Maxwell SE, & Weed K (2005). Adolescent parenting and attachment during infancy and early childhood. *Parenting*, 5(1), 91–118. doi:10.1207/s15327922par0501_4
- Maas AJB, Vreeswijk CM, & van Bakel HJ (2013). Effect of situation on mother–infant interaction. *Infant Behavior and Development*, 36(1), 42–49. Doi: 10.1016/j.infbeh.2012.10.006 [PubMed: 23261788]
- Maccoby EE (1984). Socialization and developmental change. *Child Development*, 55, 317–328. Doi: 10.2307/1129945
- Maccoby EE (2000). Parenting and its effects on children: On reading and misreading behavior genetics. *Annual Review of Psychology*, 51, 1–27. Doi: 10.1146/annurev.psych.51.1.1
- Maloney MP, & Ward MP (1976). *Psychological assessment: A conceptual approach*. Oxford: Oxford University Press.
- McCall RB (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development*, 52, 1–12. Doi: 10.2307/1129210
- McCune-Nicolich L, & Fenson L (1984). Methodological issues in studying early pretend play In Yawke TD & Pellegrini AD (Eds.), *Child's play: Developmental and applied* (pp. 81–124). Hillsdale, NJ: Erlbaum.
- McGraw KO, & Wong SP (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30–46. doi: 10.1037/1082-989X.1.1.30

- McLoyd VC, Hardaway C, & Jocson RM (2019). African American parenting In Bornstein M (Ed.), *Handbook of parenting Vol. 4: Social conditions and applied parenting* (3rd ed., pp. 57–107). New York: Routledge Doi: 10.4324/9780429398995-3
- Medoff-Cooper B, Carey WB, & McDevitt SC (1993). The Early Infancy Temperament Questionnaire. *Journal of Developmental and Behavioral Pediatrics*, 14(4), 230–235. Doi: 10.1097/00004703-199308010-00004. [PubMed: 8408665]
- Miller SA (1987). *Developmental research methods*. Englewood Cliffs, NJ: Prentice-Hall Inc.
- Mischel W (1979). On the interface of cognition and personality: Beyond the person-situation debate. *American Psychologist*, 34, 740–754. Doi: 10.1037/0003-066X.34.9.740
- Munsters NM, van Ravenswaaij H, van den Boomen C, & Kemner C (2019). Test-retest reliability of infant event related potentials evoked by faces. *Neuropsychologia*, 126, 20–26. Doi: 10.1016/j.neuropsychologia.2017.03.030 [PubMed: 28389367]
- Murry VM, Hill NE, Witherspoon D, Berkel C, & Bartz D (2015). Children in diverse social contexts In Lerner RM (Ed. In chief) & Bornstein MH & Leventhal T (Eds.), *Ecological settings and processes in developmental systems. Volume 4 of the Handbook of child psychology and developmental science* (7th ed., pp. 416–454). Hoboken, NJ: Wiley.
- Ng F, & Wang Q (2019). Asian and Asian American parenting In Bornstein MH (Ed.), *Handbook of parenting Vol. 4: Social conditions and applied parenting* (3rd ed., pp. 108–169). New York: Routledge Doi: 10.4324/9780429398995-4
- NICHD Early Child Care Research Network. (1999). Child care and mother-child interaction in the first 3 years of life. *Developmental Psychology*, 35(6), 1399–1413. Doi: 10.1037/0012-1649.35.6.1399 [PubMed: 10563730]
- NICHD Early Child Care Research Network. (2003). The NICHD study of early child care: contexts of development and developmental outcomes over the first 7 years of life In Brooks-Gunn J, Fuligni AS, & Berlin LJ (Eds.), *Early Childhood Development in the 21st Century: Profiles of Current Research Initiatives* (pp. 182–201). New York, NY: Teachers College Press.
- Nozza RJ, Miller SL, Rossman RNF, & Bond LC (1991). Reliability and validity of infant-speech discrimination-in-noise thresholds. *Journal of Speech and Hearing Research*, 34, 643–650. Doi: 10.1044/jshr.3403.643 [PubMed: 2072689]
- Nunnally JC (2017). *Psychometric theory* (3rd ed.). New York: Mc-Graw-Hill.
- Pêcheux MG, & Lécuyer R (1983). Habituation rate and free exploration tempo in 4-month-old infants. *Journal of Behavioural Development*, 6(1), 37–50. Doi: 10.1177/016502548300600103
- Pérusse D, Neale MC, Heath AC, & Eaves LJ (1994). Human parental behavior: evidence for genetic influence and potential implication for gene-culture transmission. *Behavior Genetics*, 24, 327–335. Doi: 10.1007/BF01067533 [PubMed: 7993310]
- Piaget J (1952). *The origins of intelligence in children* (Cook M, Trans.). New York, NY: WW Norton & Co Doi: 10.1037/11494-000
- Radke-Yarrow M, Zahn-Waxler C, & Chapman M (1983). Children's prosocial dispositions and behavior In Mussen Paul H. (Ed.), *Manual of child psychology* (4th ed., pp. 469–546). New York: Wiley.
- Roberts BW, & DelVecchio WF (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126, 3–25. Doi: 10.1037/0033-2909.126.1.3 [PubMed: 10668348]
- Ryan RM, & Padilla CM (2019). Transition to parenthood In Bornstein MH (Ed.), *Handbook of Parenting. Vol. 3. Being and Becoming a Parent* (3rd ed., pp. 513–555). New York: Routledge Doi: <https://www.taylorfrancis.com/books/e/9780429433214/chapters/10.4324/9780429433214-15>
- Sameroff A (2009). Designs for transactional research In Sameroff A (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 23–32). Washington, DC: American Psychological Association Doi: 10.1037/11877-002
- Saudino KJ (2012). Sources of continuity and change in activity level in early childhood. *Child Development*, 83, 266–281. doi: 10.1111/j.1467-8624.2011.01680.x [PubMed: 22103336]
- Seifer R, Sameroff AJ, Barrett LC, & Krafchuk E (1994). Infant temperament measured by multiple observations and mother report. *Child Development*, 65(5), 1478–1490. Doi: 10.1111/j.1467-8624.1994.tb00830.x [PubMed: 7982363]

- Shirley MM (1933). The first two years: A study of twenty-five babies. Volume II: Intellectual Development (Monograph series no. VII). Minneapolis, MN: University of Minnesota Press.
- Spitz RA (1965). The first year of life: A psychoanalytic study of normal and deviant development of object relations. Oxford, England: International Universities Press.
- Sroufe LA, Egeland B, Carlson E, & Collins WA (2005). Placing early attachment experiences in developmental context: The Minnesota longitudinal study In Grossmann KE, Grossmann K, & Waters E (Eds.), *Attachment from infancy to adulthood: The major longitudinal studies* (pp. 48–70). New York, NY: Guilford Publications.
- Stevenson MB, Leavitt LA, Roach MA, Chapman RS, & Miller JF (1986). Mothers' speech to their 1-year-old infants in home and laboratory settings. *Journal of Psycholinguistic Research*, 15, 451–461. Doi: 10.1007/BF01067725 [PubMed: 3772834]
- Streiner DL (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80(3), 217–222. Doi: 10.1207/S15327752JPA8003_01 [PubMed: 12763696]
- Tabachnick BG, & Fidell LS (2012). *Using multivariate statistics* (6th ed.). New York, NY: Pearson.
- Trevarthen C (1988). Universal co-operative motives: How infants begin to know the language and culture of their parents In Jahoda G & Lewis IM (Eds.), *Acquiring culture: Cross cultural studies in child development* (pp. 37–90). New York, NY, US: Croom Helm.
- van der Velde B, Haartsen R, & Kemner C (2019). Test- retest reliability of EEG network characteristics in infants. *Brain and Behavior*, e01269 Doi: 10.1002/brb3.1269 [PubMed: 30912271]
- Vandell DL (2000). Parents, peer groups, and other socializing influences. *Developmental Psychology*, 36, 699–710. Doi: 10.1037/0012-1649.36.6.699 [PubMed: 11081694]
- Weinfield NS, Ogawa JR, & Egeland B (2002). Predictability of observed mother-child interaction from preschool to middle childhood in a high-risk sample. *Child Development*, 73(2), 528–543. Doi: 10.1111/1467-8624.00422 [PubMed: 11949907]
- Weir JP, 2005 Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J. Strength Cond. Res.* 19 (1), 231–240. doi:10.1519/15184.1. [PubMed: 15705040]
- Wiggins JS (1973). *Personality and Prediction: Principles of Personality Assessment*. Reading, MA: Addison-Wesley.
- Wohlwill JF (1973). *The study of behavioral development*. New York: Academic Press.
- Yarrow MR, & Waxler CZ (1979). Observing interaction: A confrontation with methodology In Cairns RB (Ed.), *The analysis of social interactions: Methods issues and illustrations* (pp. 37–65). Hillsdale, NJ: Erlbaum.

Highlights

- Consistency in order of individuals in a group across time—*reliability*—is both psychologically meaningful and developmentally important.
- This prospective longitudinal study ($Ns = 51$ infants and mothers) examined short-term reliabilities of multiple infant behaviors and maternal parenting practices.
- The results point to small to large effect size reliabilities in infant behaviors and maternal parenting practices.
- Both infant behaviors and maternal parenting practices betray large amounts of unshared variance.
- Documenting these short-term reliabilities elucidates the nature and structure of early human dyadic development.

Table 1

Descriptive Statistics for Infant and Mother Domains and Indicators

Domains and Indicators		First Visit	Second Visit
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Infant Physical Domain		5.45 (.92)	5.48 (1.03)
	Balance	5.12 (1.28)	5.00 ^c (1.38)
	Movement	5.81 ^a (1.16)	5.90 ^b (1.21)
Infant Social Domain		.02 (.59)	-.02 (.60)
	Look at mother	.00 (.87)	.00 (1.05)
	Frequency	42.19 (17.68)	41.61 (22.22)
	Duration	177.45 (108.50)	178.83 (124.51)
	Smile	.03 (.95)	-.03 (.97)
	Frequency	10.57 (7.20)	9.82 (8.67)
	Duration	23.45 (26.93)	23.16 (23.11)
	Alert expression: duration (s)	2117.19 (340.73)	2092.32 (378.94)
Infant Exploration Domain		.02 (.70)	-.02 (.62)
	Look at objects	-.04 (.92)	.04 (.72)
	Frequency	107.65 (33.76)	111.27 (26.18)
	Duration	1101.31 (472.89)	1126.58 (461.96)
	Touch objects	.00 (.92)	.00 (.85)
	Frequency	87.99 (41.11)	77.61 (33.65)
	Duration	697.95 (317.29)	783.54 (331.45)
	Mouth objects	.18 (.96)	-.18 (.87)
	Frequency	50.94 (28.89)	39.25 (24.80)
	Duration	354.00 (207.46)	296.16 (199.00)
	Extent of exploration	.09 (.95)	-.09 (.86)
	Variety	8.86 (4.63)	8.31 (4.74)
	Density	1.65 (.87)	1.50 (.74)
	Consistency	7.69 (2.25)	7.25 (2.01)
	Efficiency of exploration	-.11 (.88)	.11 (.86)
	Variety	.72 (.17)	.76 (.19)
	Density	.54 (.16)	.60 (.17)
	Consistency	.88 (.17)	.89 (.14)
Infant Nondistress Vocalization Domain		-.03 (.96)	.03 (.83)
	Frequency	115.91 (63.11)	115.20 (51.85)
	Duration	225.13 (177.87)	248.21 (178.73)
Infant Distress Communication Domain		.09 (1.02)	-.09 (.73)
	Negative facial expression	.11 (1.12)	-.11 (.77)
	Frequency	6.98 (8.76)	5.12 (6.34)

Domains and Indicators		First Visit	Second Visit
		<i>M (SD)</i>	<i>M (SD)</i>
	Duration	34.14 (51.84)	25.69 (33.65)
	Distress vocalization	.08 (1.08)	-.08 (.83)
	Frequency	12.71 (15.76)	11.04 (13.13)
	Duration	63.38 (84.98)	48.11 (60.97)
Mother Nurture Domain		-.03 (.68)	.03 (.63)
	Feed/Burp/Wipe: duration (s)	343.12 (374.53)	399.30 (396.81)
	Bathe/Diaper/Dress/Groom/Other health needs: duration (s)	205.04 (217.67)	229.61 (299.02)
	Hold: duration (s)	852.92 (518.94)	815.49 (496.67)
Mother Physical Domain		.11 (.09)	.10 (.06)
	Physically/Verbally encourage balance	.18 (.17)	.18 (.11)
	Physical encouragement to sit	.27 (.27)	.31 (.28)
	Physical encouragement to stand	.28 (.30)	.26 (.20)
	Verbal encouragement to sit	.09 (.15)	.08 (.12)
	Verbal encouragement to stand	.10 (.18)	.06 (.09)
	Physically/Verbally encourage to move	.03 (.04)	.03 (.04)
	Physical encouragement to roll	.03 (.09)	.04 (.10)
	Physical encouragement to crawl	.00 (.03)	.01 (.04)
	Physical encouragement to step	.02 (.08)	.01 (.04)
	Verbal encouragement to roll	.08 (.15)	.09 (.15)
	Verbal encouragement to crawl	.04 (.11)	.04 (.09)
	Verbal encouragement to step	.01 (.05)	.00 (.00)
Mother Social Domain		.06 (.64)	-.06 (.67)
	Encourage attention to mother	.04 (.90)	-.04 (.93)
	Frequency	22.54 (14.38)	23.20 (14.13)
	Duration	209.22 (144.79)	181.72 (145.12)
	Social play	.06 (.89)	-.06 (.96)
	Frequency	12.45 (9.63)	11.22 (9.41)
	Duration	95.24 (86.71)	84.15 (103.69)
	Express affection	.08 (.84)	-.08 (.87)
	Frequency	23.81 (14.61)	21.25 (17.66)
	Duration	79.18 (95.54)	64.89 (65.59)
Mother Didactic Domain		-.01 (.85)	.01 (.98)
	Frequency	38.68 (23.69)	40.24 (28.07)
	Duration	374.47 (307.87)	369.73 (337.29)
Mother Material Domain		.06 (.63)	-.06 (.61)
	Quantity of objects provided	.12 (.84)	-.12 (.86)
	Variety	12.22 (5.79)	11.16 (6.08)

Domains and Indicators		First Visit	Second Visit
		<i>M (SD)</i>	<i>M (SD)</i>
	Density	3.15 (1.77)	2.67 (1.51)
	Consistency	8.63 (1.61)	8.16 (1.87)
	Quality of objects provided	.01 (.87)	-.01 (.76)
	Total responsiveness	9.11 (1.60)	9.07 (1.49)
	Number of highly responsive objects	3.10 (2.02)	2.86 (1.76)
	Percent highly responsive objects	.26 (.17)	.28 (.18)
Mother Language Domain		.11 (.91)	-.11 (.76)
	Frequency	200.07 (79.68)	175.47 (66.75)
	Duration	818.67 (457.52)	770.07 (440.03)

Note. If infants were never in the physical position necessary to exhibit a given skill during the session, no rating was made for that skill.

^a *N* = 50.

^b *N* = 46.

^c *N* = 48

Table 2

Short-Term Reliability across Time for Infant and Mother Domains

Variable	<i>N</i> ^a	<i>r</i>	
		Zero-Order	Controlled
Infants			
Physical	50	.60***	.54***
Balance	45 ^b	.74***	
Movement	48 ^b	.61***	
Social	51	.11	.10
Look at mother	51	.38**	
Smile	50	.27*	
Alert expression	51	.54***	
Exploration	49	.33**	.21 ^c .19 ^d
Look at objects	50	.45***	
Touch objects	51	.31*	
Mouth objects	51	.15	
Extent of exploration	50	.43***	
Efficiency of exploration	49	.40**	
Nondistress Vocalization	49	.08	-- ^e
Distress Communication	51	.34**	-- ^e
Negative facial expression	49	.16	
Distress vocalization	49	-.02	
Mothers			
Nurture	51	.60***	-- ^e
Feed/Burp/Wipe face or hands	50	.63***	
Bathe/Diaper/Dress/Groom/Other health needs	50	.47***	
Hold	51	.62***	
Physical	51	.37**	.36**
Physically/Verbally encourage to sit/stand	51	.36**	
Physically/Verbally encourage to roll/crawl/walk	51	-.08	
Social	51	.49***	.49***
Encourage attention to mother	51	.41**	
Social play	51	.52***	
Express affection	50	.60***	
Didactic	51	.49***	.41**
Material	50	.12	.07
Quantity of objects provided	50	.38**	

Variable	<i>N</i> ^a	<i>r</i>	
		Zero-Order	Controlled
Quality of objects provided	50	.05	
Language	51	.71 ***	-- ^e

^a Influential case(s) with Cook's D ranging from .20 to .86 were identified and removed.

^b In some cases, the infants were never in the physical position necessary to exhibit a given skill during the whole session, no rating was made for that skill for these infants.

^c Infant Exploration controlled for mother Didactic.

^d Infant Exploration controlled for mother Material.

^e Controlled correlation was not computed because there was no corresponding mother/infant behavior (infant Distress Communication and mother Nurture) or the variables were not correlated with corresponding partner's behaviors (infant Nondistress Vocalization and mother Language).

* $p < .05$

** $p < .01$

*** $p < .001$, all one-tailed tests