Reference: Waters, E. & Deane, K. (1985). Defining and assessing individual differences in attachment relationships: Q-methodology and the organization of behavior in infancy and early childhood. In I. Bretherton & E. Waters (Eds.), Monographs of the Society for Research in Child Development, 50, nos. 1-2. pp. 41-65.

DEFINING AND ASSESSING INDIVIDUAL DIFFERENCES IN ATTACHMENT RELATIONSHIPS: Q-METHODOLOGY AND THE ORGANIZATION OF BEHAVIOR IN INFANCY AND EARLY CHILDHOOD

EVERETT WATERS AND KATHLEEN E. DEANE

State University of New York at Stony Brook

At times, it seems as if attachment research could fall victim to its own success. In the span of barely 15 years, we have come to accept Freud's view that attachment in infancy constitutes a genuine love relationship. We have recognized that this relationship is closely tracked by patterns of behavior toward caregivers and that this behavior is complexly organized, goal-corrected, and sensitive to environmental input. We have also adapted observational techniques employed by behavioral biologists and learned to examine infant behavior in detail and in context. As a result we have learned a great deal about attachment and exploratory behavior and about the organization and motivation of proximity seeking, contact maintaining, avoidance, and resistant behavior during separation and reunion.

As a direct result of these accomplishments, we have been able to standardize and validate useful techniques for assessing individual differences in attachment behavior and in the infant's ability to use an adult as a secure base from which to explore. These tools have enabled us to outline the antecedents of individual differences in attachment behavior and to establish that attachment relationships can be markedly stable over significant periods of time. At the same time, we have been able to rule out both intelligence and temperament as alternative explanations of individual differences in secure versus anxious attachment and to demonstrate that patterns of attachment can change in response to significant changes in patterns of care. Evidence that infants who are secure with one parent are not necessarily secure with the other is particularly decisive on the distinction between attachment and temperament and confirms that our assessments reflect characteristics of specific relationships rather than traits of particular infants.

In the midst of these and other advances that are well documented in this *Monograph*, it is easy to lose sight of the fact that there have been very few non-laboratory observations of attachment behavior during the last 10 years. Reports on attachment behavior outside the 12-18 month age range have also been few and far between. Questions about what is learned during the formation of attachment relationships, about the course of attachment after infancy, and about individual differences beyond security and anxiety have received surprisingly little attention. Unfortunately, the longer these questions are left unanswered, the more difficult it becomes to design incisive research or to assimilate new data. In fact, most of the recent data on the correlates of secure versus anxious attachment are simply being assimilated to the general hypothesis that "all good things go together." This provides little guidance for further research and almost insures that each new study will be less and less incisive. It also introduces the risk that the attachment construct will lose its definition and once again fall in among the feckless personality trait variables from which it was only recently rescued.

This research was funded in part by a grant from the Foundation for Child Development (Program for Young Scholars in Social and Affective Development). The authors wish to thank the many colleagues who contributed their expertise to this project by performing the criterion Q-sorts for security, dependency, sociability, and social desirability.

ATTACHMENT AND THE SECURE BASE PHENOMENON

The cornerstone of Bowlby's attachment theory was replacement of psychoanalytic drive reduction theory with a control systems analysis. In Bowlby's view, the apparently purposive behavior of infants toward caregivers can be explained in terms of a behavioral control system, which functions to maintain a balance between attachment and explora-

Note: Monograph chapters used a combined bibliography. See original volume for references.

tory behavior across a wide range of contexts. In familiar contexts, and in the absence of what Bowlby called "natural cues to danger," the balance favors exploration punctuated by periodic checks on the adult's location. In other contexts, which may have entailed risk of injury or predation in the environment to which humans are adapted or which may have been associated with negative consequences in a particular infant's experience, the balance favors physical contact over exploration. When the control system operates as designed, it enables the infant to play an active role in its own behavior and development and facilitates both social and cognitive development. The operation of an attachment control system over time is referred to as the secure base phenomenon.

When Bowlby (1969/ 1982b) introduced behavioral control systems into attachment theory, he was addressing a very specific problem related to the motivational model underlying attachment theory. Aside from this, his interest was primarily in attachment as an emotional bond. Since the publication of the first volume of *Attachment and Loss,* many elaborations and alternatives to the emotional bond concept have been proposed. The infant-adult bond has been variously conceptualized as a traitlike variable (e.g., Coates, Anderson, & Hartup, 1972), a response class (e.g., Masters & Wellman, 1974), a relationship (e.g., Hinde, 1976, 1979), one facet of a broader social network (Weinraub, Brooks, & Lewis, 1977), and an organizational construct (Sroufe & Waters, 1977).

Each of these proposals conveys something important and well worth incorporating into assessment and empirical research. But unfortunately, insights and alternatives have rarely been presented in terms that are easily translated into assessment procedures. This poses several problems. First, theories that are not readily translated into assessment procedures are not easily tested. Second, a test or task cannot be made relevant to attachment theory by fiat. Unless a procedure is carefully tailored to assess attachment as it is defined in theory, it can only generate irrelevant and unassimilable data. This is hard to achieve when attachment is defined in psychodynamic, organismic, or contextual terms. But when attachment is closely tied to the performance of a behavioral control system, a clear relationship to patterns of secure base behavior in the home is easily defended as the criterion against which the construct validity of any attachment measure has to be assessed.

While Ainsworth clearly shares Bowlby's view that attachment is an emotional bond, she has come closer than anyone else to equating attachment per se with the control system that organizes attachment behavior. In her early cross-cultural work (Ainsworth, 1967) and in her Baltimore longitudinal study (Ainsworth et al., 1971), she detailed the infant's use of an adult as a secure base from which to explore. In addition she proposed that secure base behavior across time and across situations, rather than separation protest per se, is the most decisive evidence of attachment in infancy (Ainsworth, 1973). Finally, she has conceptualized secure versus anxious attachment in terms of differences in the infant's ability to use an adult as a secure base in naturalistic settings (Ainsworth et al., 1971).

When attachment is identified with a behavioral control system, specific behavioral referents are built into the definition of the construct. As a result, assessment problems become more tractable, and it is easier to say exactly what the development of attachment is supposed to be the development of. In addition, emphasis on a behavioral control system is entirely consistent with the notion that attachment is learned (Waters & Deane, 1982). Once we are past this perennial stumbling block, we can move directly to more significant issues: what is learned in the acquisition of a first attachment, what role (if any) do species-specific biases in learning ability play in the acquisition of attachment, and what is learned in forming a first attachment that facilitates acquisition of attachment to additional caregivers?

A second advantage of identifying attachment with an underlying control system is that the performance of a control system is easier to assess than a relationship, an organizational construct, or a social network. In addition, reference to a control system and to patterns of secure base behavior suggests criteria against which various approaches to assessment can be evaluated. This is most important because, as we have indicated in several recent papers (e.g., Waters & Deane, 1982), incongruous weddings of theory and assessment have led to fruitless controversy and wasted effort. In addition, they have been extremely difficult to unravel (Waters & Deane, 1982).

The primary disadvantages seem to be that the control system/secure base perspective is not as relevant to describing or explaining responses to loss of attachment figures as we might like and that there is the risk that we may reify the behavioral system, only to discover later that it is more useful as a

metaphor than as an explanation for attachment behavior. At the same time, our inability to assess attachment beyond a limited age range is becoming a more acute problem, and the trend away from research on attachment per se toward the study of attachment correlates continues. Hence, for the moment, the advantages in tying the attachment construct closely to an underlying behavioral system and to the secure base phenomenon seem to outweigh the disadvantages.

CRITERIA AGAINST WHICH ASSESSMENT PROCEDURES CAN BE EVALUATED

When attachment is conceptualized in terms of a behavioral control system and closely tied to the secure base phenomenon, we can define seven criteria against which any measure can be evaluated. Where measures are not available or where conventional measures fall short, the same criteria can serve as guidelines in the development of new measures.

Structural Fidelity

In her classic monograph on the relationship between psychological assessment and psychological theory, Loevinger (1957) outlined a number of important conditions for valid assessment. Among these she included the notion that a measure should provide data congruent with the type of construct it is designed to assess. Quantitative traits call for continuous variables; taxonomic entities call for classification rather than measurement; multiple components call for multidimensional assessment. When theory and measurement are congruent in this respect, the measure affords structural fidelity.

The defining characteristic of the attachment system is reference to a set-goal. That is, the measure of a control system should be in terms not of gross behavioral output but of success or failure at some regulatory function (i.e., adaptiveness). This involves assessment across time and across the range of stimuli to which the system is responsive. As long as the distinction between (quantitative) output and (qualitative) performance is maintained, adaptiveness can certainly be measured in terms of a continuous variable.

Reference to Behavior

The control systems described by Ainsworth et al. (1971), Bischof (1975), Bowlby (1969/ 1982b), Bretherton (1980), and Waters and Deane (1982) all incorporate, inter alia, sensitivity to stimuli in the social environment. In addition, each of these control systems generates observable behavioral output. Thus measures of individual differences in attachment should make specific reference to behavior. This might involve defining behavioral variables as narrow as "clinging to mother's leg" or as broad as "contact maintaining" (which might include reaching, clinging, resisting being put down, etc.).

Once these primary data have been collected via time sampling, rating, or some other method, it may be economical to describe the performance of the attachment behavioral system in terms of trait descriptive adjectives. For example, it is more economical to describe a child as "secure" than it is to say that the child explores and approaches novelty more readily when the adult is present, monitors adult location and behavior spontaneously and effectively, is more tolerant of self-initiated separation than of adult moving away, retreats to adult when distressed, is not angry after brief separation, finds physical contact a potent stimulus for terminating distress, prefers to be comforted by this adult, et cetera. The essential point is that an attachment measure should provide primary data that have clear behavioral referents. Trait language should be used only to summarize behavior-never as a substitute; never as an explanation.

Take the Context of Behavior into Account

The control systems described by Ainsworth et al. (1971), Bischof (1975), Bowlby (1969/ 1982b), Bretherton (1980), and Waters and Deane (1982) all make specific reference to the context in which attachment behavior occurs. This involves both the context provided by objects and events in the environment and the context provided by the infant's or child's ongoing behavior. For example, approaches to the mother may be phenotypically similar; but, within a control systems perspective, some are "exploratory approaches" (i.e., not attachment behavior), and some are "proximity seeking" or even a "retreat" from a stranger (i.e., cornerstones of the secure base phenomenon). Similarly, an incomplete approach in a nonseparation context has to be distinguished from an incomplete approach when the caregiver returns after a separation. Such distinctions have to be made because phenotypically similar behaviors often have dramatically different external correlates when the context in which behavior occurs is taken into account (e.g., Hay, 1980; Sroufe & Waters, 1977; Tracy, Lamb, & Ainsworth, 1976; Waters, Matas, & Sroufe, 1975).

The emphasis that control systems theory places on the context in which behavior occurs has

two important implications for assessment. First, an attachment measure should explicitly recognize that a given behavior can be congruent with the system's set-goal in one context and incongruent, even irrelevant to the attachment construct, in other contexts. Second, a measure should preserve or incorporate information that allows us to score different behaviors as equivalent when their outcome vis a vis the control system's set goal is equivalent.

Evaluate Relationships among Affect, Cognition, and Behavior

Each of the control system models mentioned above includes sensitivity to affective, cognitive, and behavioral cues. The primary effect of affective inputs seems to be on the set-goal of the attachment behavioral system. Infants are more tolerant of separation and explore farther away when they are alert and not distressed. They are less tolerant of separation and generally demand more physical contact when they are ill, hungry, concerned by recent experiences, et cetera.

The primary role of cognitive and behavioral input to the control system is regulatory. Cognition plays a critical role in monitoring adult behavior and accessibility, in recognizing and evaluating events in the environment, and in exploratory interactions with toys and people. Finally, behavior is the attachment control system's effector mechanism. It is used to correct deviations from the set-goal by changing the proximity of the infant to the caregiver or by stimulating changes in the caregiver's behavior.

All these are monitored and integrated in order to keep the balance between contact and exploration within the limits defined by the control system's set-goal. None of them has any decisive meaning to the control system, except in the context of the others. Accordingly, any attachment assessment should include data from all three domains.

Allow for Nonquantitative Developmental Change

There are many ways in which the operation and output of a control system can change during development. Some of the more important among these include changes in the system's set-goal, in the inputs to which the system is sensitive, in the integration of system components, in the behaviors through which adaptive response is effected, and in the relationship between the system in question and other behavioral systems. Each of these involves the structure and configuration of the control system and may or may not alter the rate of any particular behavior. Accordingly, an attachment measure should anticipate the problem of detecting and describing change in how the behavioral system operates. Emphasis on age changes in the frequency of particular behaviors will not suffice.

Detect Coherence over Time Even in the Context of Behavioral Change

As described above, an attachment measure that offers structural fidelity will employ a criterion of "adaptiveness" (i.e., how well is the system tracking its set-goal?) rather than assess the quantity of behavioral output per se. While both criteria involve assessment over time, they refer primarily to cross-sectional assessment. A related criterion is that a measure should be able to detect consistency (if any) in adaptive functioning across age, even when there are developmental changes in the behavior through which adaptive response is effected.

An Additional Criterion: Discriminant Validation

The behaviors initiated, modulated, or terminated by an attachment control system are not properly part of the system. They belong to the infant's or child's repertoire of action skills. Thus the phenotypically similar behaviors can be performed in the service of attachment in one context and be unrelated to the attachment control system's activity in another. This raises two problems that psychometricians have discussed in terms of discriminant validity. First, data collected to assess attachment may be subject to alternative interpretations, especially in terms of temperament, intelligence, and other trait constructs. The most decisive approach to this is to incorporate assessment of these variables into the procedure for assessing attachment. This enables us to evaluate and rule out plausible alternative interpretations. A second problem arises from the fact that evaluative biases can easily intrude into observational data (e.g., the tendency to attribute socially desirable secure base behavior to preferred subjects).

One response to this is to anticipate it and to employ assessment strategies and observational designs that minimize or attenuate observer biases. An additional strategy is to include specific assessment of social desirability bias in the measure. Discriminant assessment is an important component of any individual differences assessment, regardless of how attachment is conceptualized.

EVALUATION OF CONVENTIONAL ASSESSMENT METHODS

When trait ratings, time sampling, and the widely used Strange Situation procedure are evaluated in terms of these criteria, each shows a unique pattern of assets and liabilities. A brief review of these. conventional assessment methods can be quite useful since even methods that fail to meet one or more of the criteria mentioned above are likely to have particular strengths that should be incorporated into some other method to provide better assessment.

Trait Ratings

The primary advantages of rating methods are flexibility and economy. In terms of the criteria outlined above, rating methods excel in their ability to take the context in which behavior occurs into account and in their sensitivity to coherence over time, in the context of underlying behavioral change. These advantages arise from the ability of (some) human judges to apply complex cognitions in the task of summarizing and of scaling diverse behavioral inputs in terms of a theoretically defined criterion such as security. (See Meehl, 1973, for detailed consideration of the conditions under which clinical judgment is both useful and economical.) In principle, rating methods are quite amenable to discriminant validation, though it is rarely undertaken with the care it deserves.

The primary disadvantages of rating methods are that they offer little in the way of structural fidelity since attachment is not a trait, that when primary data are collected in terms of ratings it is impossible to recover the behavioral details on which the ratings were based, and that ratings are exceptionally susceptible to intrusion of biases, response sets, and global rating styles. As a result, ratings are coercive; they force us to see the phenomenon in terms of constructs built into the rating scale and limit our ability to construct and evaluate alternatives without new data. Rating methods also tend to be conservative. The set adopted in rating a particular construct works against the adoption of new perspectives during observation.

Rating methods are not well suited to analyses of nonquantitative developmental change, even though they can enable us to see beyond age changes in specific behaviors in order to detect continuity of adaptation. In particular, rating data do not provide the behavioral detail necessary to analyze developmental changes in a behavioral system's set-goal, the inputs to which a system is sensitive, or relationships among system components.

Finally, ratings generally incorporate information about affect, cognition, and behavior. Thus they can be quite useful if the alternative is a method that tends to overlook any of these domains. At the same time, it is inherently difficult to separate affect, cognition, and behavioral variance in rating data. Moreover, in most contexts, a process level analysis of affect-cognition-behavior interplay is much more valuable than approaches that emphasize correlations and components of summary score variance without respect to the sequencing of events in real time.

Observational Data: Frequency Counts and Time Sampling

The primary advantage of frequency counts and time sampling procedures is that they retain much of the behavioral detail that is sacrificed in rating methods. At the same time, they are extremely expensive to employ. In terms of the criteria outlined above, observational data can afford considerable structural fidelity if the categories of observation are tailored to the design of the behavioral system and take the context in which behavior occurs into account. In practice, most observational scoring schemes are extremely insensitive. They tend to aggregate broad ranges of phenotypically similar behavior, even where the same behavior has markedly different implications for control system functioning in different situational or behavioral contexts (e.g., Coates, Anderson, & Hartup, 1972).

Individual differences in observational data across time have traditionally been analyzed in terms of stability within particular behavior categories (e.g., Masters & Wellman, 1974). Unfortunately, the stability of specific behaviors clearly underestimates the coherence and consistency of adaptive functioning across periods of significant behavioral change (Waters & Sroufe, 1983). In principle, we can define behavior categories that have clear implications for adaptive functioning at various ages and intercorrelate scores on age appropriate measures across time intervals. Alternatively, we can employ observational protocols that are sufficiently detailed to support secondary data reduction, in which frequency counts or time sampling data are reduced to ratings.

Unfortunately, very little of the potential inherent in observational data has been realized in developmental research. Several problems have contrib-

uted to this. First among these is the lack of conceptual analysis and the relative unfamiliarity with the target behaviors reflected in most lists of observation categories. Aside from this, time sampling and frequency counting tax the skill, patience, and endurance of even the most dedicated researcher. In addition, the number of behavior categories that can be assessed at once (or in one viewing of a video record) is quite small while at the same time the number of behavior categories increases dramatically when distinctions among superficially similar behaviors are made and when the context in which behavior occurs is taken into account. Finally, many behaviors that are of interest in developmental research occur at low and/or uneven frequencies. This poses serious problems for the collection of psychometrically reliable data unless very substantial periods of observation are devoted to each subject (Waters, 1978).

The Ainsworth Strange Situation Procedure

The Ainsworth Strange Situation procedure involves both a standardized observation context and a set of scoring protocols. Both the observation context and the scoring system were developed with explicit reference to an attachment behavioral control system. The separation-reunion episodes of the Strange Situation were designed to reproduce mild to moderately stressful challenges to the attachment behavioral system that occur in everyday situations. The system for scoring interactive behavior in the Strange Situation is explicitly designed to evaluate the adaptive functioning of the attachment control system. That is, it parses behavior in terms of inputs and contexts that are theoretically relevant to the control system's task of maintaining an optimal balance between proximity and exploration. Accordingly, the assessment affords exceptional structural fidelity.

In research on individual differences, the Strange Situation procedure is more a psychometric instrument than an observational measure because the behavior observed in the laboratory is not, in and of itself, the behavior of interest. That is, crying in the Strange Situation is not assessed as a sample of crying rate in the home, nor are rates of proximity seeking, avoidant behavior, or resistant behavior in the laboratory expected to correlate with rates of similar behavior at home. These behaviors are assessed in the Strange Situation as signs or predictors of the control system's ability to operate over a wide range of contexts and to organize behavior toward the adult over significant periods of time.

The validity of Strange Situation assessment depends on its ability to substitute for extensive assessment of the secure base phenomenon in naturalistic contexts. The primary evidence in support of the procedure's external validity is detailed in Ainsworth et al. (1978). If the relationship between Strange Situation classifications and the secure base phenomenon at home was not evident in empirical data, the procedure might still provide valuable data on the quality of attachment relationships. But we would be unable to explain why or how.

The principal limitations of the Strange Situation procedure are that it is only applicable within a narrow age range (perhaps as narrow as 12-18 months), that repeated assessments have to be spaced to prevent strong carryover effects, and that the situation and scoring procedures do not lend themselves to research on developmental changes in the attachment control system. The procedure is also expensive to administer and score, and scoring is difficult to learn without direct instruction.

One additional aspect of the Strange Situation procedure seems problematic. The scoring system that is best validated for research on individual differences involves taxonomic classification rather than quantitative assessment. Moreover, the distribution of subjects across the avoidant, secure, and resistant categories is markedly unbalanced. As a result, the data gleaned from the procedure offer less information and fewer options for data analysis and require larger samples than either rating or time sampling methods. It can be argued, however, that taxonomic assessment is part and parcel of the procedure's structural fidelity, which may explain why the classification system has had markedly greater external validity than any other scheme for scoring Strange Situation data.

Summary

Each of the conventional methods for assessing attachment behavior has a distinct pattern of strengths and weaknesses. Ratings are economical and flexible, but they do not provide behavioral detail or structural fidelity. They are coercive and conservative. Observational methods preserve behavioral detail, but they tend to become unmanageable when we try to take the behavioral context into account. In addition, observational methods are rarely informed by theory or pilot observation and thus often fail to deliver on much of their potential.

The Strange Situation procedure affords well-validated assessment, but it is applicable only to a

very limited age range. It is not well suited to research on important issues in attachment theory that call for multiple assessments or for assessment of developmental change. The procedure is also difficult to score and yields taxonomic rather than quantitative data.

In brief, none of the conventional methods for assessing attachment meet all the criteria outlined above. Thus there is a compelling case for developing alternative measurement strategies.

THE Q-SORT METHOD: AN ALTERNATIVE APPROACH TO ATTACHMENT ASSESSMENT

The Q-sort method was introduced by Stephenson (1953) and has been used extensively in personality assessment and developmental research (e.g., Baumrind, 1968; Bem & Funder, 1978; Block, 1961/1978; Block & Block, 1980; Roberts, Block, & Block, 1984; Waters, Garber, Gornal, & Vaughn, 1983). Q-sort methodology consists of three components: procedures for developing sets of descriptive items to which scores are to be assigned; procedures assigning scores to items by sorting them into a rank order, from most characteristic to most uncharacteristic within each subject; and a wide variety of procedures for data reduction and analysis.

During the last 2 years we have developed and pilot tested a 100-item Q-set that allows us to meet each of the criteria for attachment assessment outlined above and at the same time affords the psychometric and data analytic advantages that are unique to the Q-sort method. As is not the case with conventional trait assessment Q-sets, each item in this Q-set makes specific reference to behavior. Many of the items qualify their behavioral referents by specifying a specific context. The Attachment Q-set covers a broad range of secure base and exploratory behavior, affective response, social referencing, and other aspects of social cognition. Accordingly, the Q-set can be construed as an overview of the entire domain of attachment relevant behavior, as currently understood within an ethological/control systems perspective.

The Q-set: A Vocabulary for Describing Attachment Behavior

Each item in the attachment behavior Q-set consists of an item title and more specific descriptive statements printed individually on cards. These items constitute a standard vocabulary for describing individual differences within a particular domain of personality, attitudes, or behavior. Q-set items can easily be written to refer to specific behaviors or to behavior in specific contexts. In addition to describing individuals, Q-sets can also be used to operationalize constructs in terms of an array of scores on a specific set of relevant items.

Use of a standard multiple-item vocabulary for assessment has many advantages. The development of the Q-set itself demands close examination of theory and reference to extensive clinical or observational data. What is the range of behavior relevant to a particular construct or set of constructs that might be assessed with the Q-set? What contexts are salient in evaluating construct relevant behavior? Construction of a Q-set forces us to consider process-oriented models of behavior and behavioral organization in detail. We are also forced to clarify distinctions and ambiguities that are more easily glossed over in designing rating scales. A well-designed Q-set is thus a powerful tool for transferring theoretical and behavior sophistication to new observers.

Use of a standard Q-set insures that diverse observers evaluate the same content in describing each subject. The use of a broad-band item set is also economical. After individuals have been described in terms of a particular Q-set, a wide range of variables can be scored from this description, including variables that attracted attention only after data collection was completed (see Data Reduction and Analytic Strategies below).

Q-Sorting Procedure

When judges or observers use a Q-set to describe a subject, they sort the items (on cards) into piles whose designations range from most characteristic to least characteristic of a particular subject. This is usually accomplished in several steps, by sorting the items into three piles and then subdividing these into a total of nine. Then, working from the outer piles toward the center, each pile is adjusted so that the final sort conforms to a symmetrical, unimodal distribution with specified numbers of items in each of the nine piles (i.e., 5, 8, 12, 16, 18, 16, 12, 8, and 5).

Each item is scored in terms of its placement (piles 1-9) in the distribution (e.g., each of the five items in pile 9 receives a score of 9, each of the five items in pile 1 receives a score of 1, and so forth). When several sorters describe the same subject, a composite description can be constructed by averaging the scores assigned by each sorter to each item.

The primary advantages of the Q-sort method are that observers can be kept unaware of the constructs that will be scored from the data they provide; that observers are not required to have detailed knowledge of norms for each item, as they are for conventional rating methods; that response biases are reduced by sorting items into a fixed distribution; that the significance of a behavior is clearly distinguished from the frequency with which it occurs; that each item is explicitly scored in the context of a well defined set of other items; and that data from different samples can be compared directly because sample norms do not enter into the scoring. In addition, description of subjects in terms of an array of scores on items with highly specific content affords a wide range of analytic possibilities that are not available when rating procedures are employed to summarize a wide range of information in a single score.

Data Reduction and Analytic Strategies

As mentioned above, the Q-sort method involves describing a subject in terms of an array of scores on a standard set of items. One of the greatest advantages of this approach is the wide range of data reduction and analytic strategies that can be employed in such data.

Reliability assessment. Agreement between independent observers provides evidence that they are performing their tasks similarly (i.e., that they are in some sense interchangeable). This is important when one observer is very experienced or can otherwise be considered the criterion against which correct scoring should be assessed and when the observations of different subjects are undertaken by different observers.

Unfortunately, agreement per se does not decisively establish that observations are accurate. It merely suggests that observers have performed similarly. More important, agreement does not imply that the observations on which observers agree are typical or representative of the subjects' mean rates or relative performances. This is especially problematic when data are based on very brief encounters with a subject, as in many field studies and most laboratory procedures.

The psychometric reliability of a measure reflects the extent to which the scores assigned to individual subjects are representative of their typical or "true" scores over the entire set of comparable occasions that might have been assessed. When data are unreliable, correlations among scores are attenuated and the statistical power of group comparisons is diminished. In most developmental research, observational data tend to be extremely accurate but may be unrepresentative (when this distinction is not made, reliability in the sense of representativeness is generally not evaluated).

In Q-sort data, agreement can be assessed in terms of the reliability of individual items or of an entire sort. Item agreement is assessed by intercorrelating the scores assigned by two sorters across a sample of subjects. Agreement on a complete sort can be assessed within each subject by intercorrelating the arrays of scores assigned to a particular subject by two sorters, and mean agreement can be assessed by computing the average agreement across all the subjects in a sample.

In general, it is extremely useful to have several sorters describe each subject in a study because the reliability of a Q-sort description increases when several sorts are averaged to obtain a composite Q-sort description. This is simply an instance of the well-known relationship between test length (each Q-sort description being considered a single item) and reliability (Ghiselli, Campbell, & Zedeck, 1981) and of the notion that the more points of view and the more observational occasions included in a description, the more representative the description will be (Rushton, Brainerd, & Pressley, 1983).

When several Q-sorts are obtained to describe a single subject or a single hypothetical construct, the reliability of a composite of these sorts can be assessed by computing the mean intercorrelation among the scores that several expert raters assign on a particular item across all the subjects in a sample and then applying the Spearman-Brown formula (Ghiselli et al., 1981). In addition, since each subject is described in terms of an array of scores for the complete Q-set, the reliability of the full Q-sort description of any individual subject can be obtained by intercorrelating the arrays of scores assigned to the subject by several observers and applying the same Spearman-Brown formula.

The significance of these reliability coefficients is that they indicate the proportion of variance in a set of data that is reliable (as opposed to indicating error), that they can be used to determine the degree to which correlations and the power of statistical tests will be attenuated as a consequence of error variance in the data, and that they can be used to determine how much observation time would be necessary to obtain reliable Q-sort descriptions. Psychometric reliability estimates can also be computed

for rating data and observations (e.g., Waters, 1978), though they are rarely reported in developmental research. A special advantage of Q-sort data is that reliability coefficients can be computed for individual subjects as well as for a sample at large. By indicating when additional data should be collected on a particular subject, this provides a useful approach to quality control in Q-sort data.

Analysis of individual items. The most obvious approach to Q-sort data is simply to treat the scores assigned to subjects on the Q-set items as separate scores on so many different variables. These scores can be correlated with other data, or groups can be compared in terms of item-by-item significance tests (e.g., Vaughn, Deane, & Waters, in this vol.; Waters et al., 1983; Waters et al., 1979). Since this typically involves a large number of statistical tests, it is important to deal with the fact that a certain number of significant results will occur by chance. One approach is simply to employ conservative significance criteria. An alternative is to undertake a cluster analysis of the variables on which significant results are obtained and to emphasize domains of significant effects rather than specific items in interpretations and discussion.

Subsets of items as scales. A wide variety of procedures can be used to compose scales from selected Q-set items after a complete Q-set has been sorted. These range from using rational criteria to select items on the basis of specific content to using cluster analysis or factor analysis for empirical item analysis. The full range of item-weighting procedures can also be employed. When the scores assigned to a selected subset of items are summed or weighted and summed, the total score can be used as an index of a single construct. The reliability of such a score can be determined from the mean intercorrelations among the items included in the scale (Ghiselli, Campbell, & Zedeck, 1981).

Criterion sorting. Judges can use a Q-set to operationalize important attachment constructs by sorting the items to describe a hypothetically most secure, dependent, or sociable subject. Item-by-item comparisons between the placement of items by sorters defining one construct and sorters using the same items to define a different construct can be used to evaluate similarities and differences among related constructs (e.g., Deane & Waters, 1984; Waters, Noyes, Vaughn, & Ricks, in press).

When constructs have been defined in this way, subjects can be scored on each construct by computing the correlation between the composite description of the "hypothetically most x subject" and the Q-sort description of a particular subject (i. e., the correlation between two arrays of scores within each subject). The correlation coefficient between the construct definition and the description of the subject is used as the subject's score on that construct. The more similar the subject is to the hypothetically most extreme subject, the higher the subject's score on the construct. This procedure has several important advantages. First, it enables us to place some distance between the observers who collect the primary data in a study and the constructs that will be scored from their data. Biases and halo effects are much less likely to intrude when observers use a Q-set to "describe a subject's behavior" without reference to any specific constructs than when observers are asked to assign ratings on the constructs themselves. Second, it allows us to employ experts' definitions of a construct to score subjects without having to enlist the experts as observers.

The criterion-sort procedure also insures that the full range of relevant behavior is considered in assigning scores on a construct. Most constructs have implications for a wide range of behaviors. And in principle a high score should be reserved for subjects who have a broad profile of construct relevant behavior. In practice, raters who have a particular construct in mind respond strongly to positive evidence (e.g., to a few clear signs of insecurity in a particular interaction). As a result, moderate to high scores are often assigned to subjects whose behavior is unexceptional in much of the domain relevant to the construct. In contrast, a Q-sorter's task is to describe the subject's behavior with equal attention to every Q-set item.

After the subject has been described in detail, a high correlation between this description and a criterion sort implies exceptional behavior across a significant range of construct-relevant behavior. Isolated events are much less likely to result in high scores in Q-sort data than they would be with conventional rating methods.

Finally, this procedure enables us to develop criterion sorts and assign scores on new variables long after data collection has been completed. This is a great advantage in longitudinal research. Interpreting unexpected results or alternative hypotheses is often facilitated by the ability to score subjects on a variable for which no specific measure was included in the study.

Cluster analysis of subjects. In many research contexts, it is useful to identify homogeneous subsets of subjects within a sample. In some cases, taxonomic analysis of the sample is significant in itself. In other cases, this is merely a step toward analysis of a hypothesis that may fit a subset of the sample better than the sample as a whole. Q-sort data lend themselves to this kind of analysis quite readily because intercorrelations among subjects (sometimes called Q-correlations) can be computed across selected items or ' across an entire Q-set. These correlations reflect the similarities and differences among subjects and can easily be employed in cluster analyses of subjects.

Once homogeneous subsets of a sample are identified, they can be characterized in terms of mean profiles on individual Q-set items. Comparisons among the subsets can also be made in terms of mean scores on Q-set items or group differences on data from other sources via discriminant analysis. Once subsets of a sample have been adequately characterized, etiological hypotheses, differential predictive hypotheses, et cetera can be made with reference to specific groups. Block (1971) and Block (1969a) have demonstrated the power of this type of data analysis in several longitudinal studies of personality development.

An Attachment Behavior Q-Set for Infancy and Early Childhood

The Attachment Q-set was developed and revised in four stages. In the first stage we reviewed the literature on attachment theory and attachment behavior in American and other samples. In the course of this review, we compiled a list of behaviors and contexts that were mentioned in theoretical articles or empirical research. In addition, we developed a list of constructs that one might want to score from a well-designed Attachment Q-set. These included security, dependency, detachment, self-efficacy, several aspects of object orientation, communication skills, predominant mood, response to physical comforting, fearfulness, anger, and trust. In a series of home visits, we rated infants and toddlers on each of these variables and subsequently specified the particular behaviors that had led to or seemed congruent with these ratings. These, plus the behaviors mentioned in the literature, constituted a preliminary item set. We then listed each of our preliminary items on cards and defined behavioral responses that would be considered the opposite of each item.

In a second series of home visits, we used the preliminary item set to describe the behavior of infants and toddlers. Behaviors that never occurred, could not be sorted with good agreement, or had very little variance across subjects were revised or eliminated. In addition, a number of items were qualified, and in several instances a single item was replaced with two or more items in order to distinguish among theoretically significant contexts.

In the third stage of development, we categorized the entire set of items and opposites and then eliminated either the item or its opposite in order to balance social desirability within each category. At this point the item set consisted of 100 items. The content categories and the number of items per category are presented in Table 1.

TABLE 1

ATTACHMENT Q-SET:

CONTENT CLUSTERS

Cluster Description	Items (N)
1. Attachment behavioral system:	
Attachment/exploration balance	12
Differential response to parents	9
2. Affectivity	19
3. Social interaction	18
4. Object manipulation	14
5. Independence/dependency	14
6. Social perceptiveness	8
7. Endurance/resiliency	6
Total	100

In the final stage of development, we asked parents to familiarize themselves with the items and use them to describe their own infant's or toddler's behavior. Two observers visited each of the subjects for 3-4 hours at home on two occasions and used the items to describe the subjects' behavior. Items on which it was difficult to obtain agreement between observers were clarified. A number of items were restated to eliminate technical terms that the mothers asked to have defined (further effort is still needed in this direction). Finally, items for which the opposite behavior (i.e., the meaning of low placement in the sort) was unclear were revised. The opposite of the behavior described in these items was added in italics at the bottom of the card on which the item is printed.

In the present item set, each item consists of a title and a definition that refers to specific behaviors and relevant contexts, and opposites are defined in italics as needed. The item titles for the Attachment Q-set are listed by item number in the Appendix 1

The Attachment Q-Set in Naturalistic Contexts

In a recent study, we visited 50 3-year-olds in their homes and collected Q-sort descriptions of each child's behavior from observers as well as from the child's mother (Deane & Waters, 1984). Three visits were scheduled at the family's convenience. On the first occasion, Observer 1 visited for 3-4 hours, accompanying the parent and child throughout the home, around the yard and neighborhood, and on any excursions away from home. The observer was responsive to bids for interaction. On the second occasion, both Observer 1 and Observer 2 visited the family for 3-4 hours. The purpose of a joint visit was to allow the now familiar observer to introduce the second observer to the family. The third visit was made by Observer 2 alone. Both graduate students and undergraduates served as observers.

Note: The complete items and sorting instructions are available on-line at www.johnbowlby.com.

After completing two visits, the observers sorted the items as described above. Correlations between the two Q-sort descriptions of each child ranged from .75 to .95. Thus 6-8 hours over two occasions provides sufficient information for sorters to provide highly reliable data.

When the series of visits with a family was completed, the procedure for providing a complete Q-sort was explained to the mother, and she was asked to familiarize herself with the Q-set by sorting the cards into three piles (characteristic, neither characteristic nor uncharacteristic, and uncharacteristic). She was encouraged to ask about both the items and the sorting procedure.

During the following week, the mother observed her child with the intent of providing a Q-sort description at the end of the week. During the next week, she observed again and completed a second sort. These two sorts were averaged together to yield a composite description of the child. The correlation between the composite mother's sort and a similar composite of the observers' descriptions of her child ranged from .59 to .93. The mean correlation was .80. Examination of differences between mothers' and observers' sorts indicated that, in many instances, the differences were clearly examples of the mother having better access to the behavior than the observers did.

These results provide clear evidence that mothers can provide exceptional data on their children's attachment behavior when they are informed in advance of what they should observe and how it is to be reported and when the procedure for reporting involves non-evaluatively stated items and a forced choice procedure such as a Q-sort.

We are currently in the midst of a similar study of 1-year-olds. While the Q-set items seem equally relevant at this age, it appears that three visits may be more appropriate than two. The items do not seem appropriate for infants under 10 months.

Criterion Sorts: Security, Dependency, and Sociability

In addition to our observational studies, we have used the Attachment Q-set in a study of security and related concepts as they apply to children aged 12 months and 3 years (Deane & Waters, 1984). Forty-three Ph.D. psychologists familiar with developmental theories of security, dependency, or sociability provided Q-sort definitions of these constructs. Eight independent sorts were collected for each construct as it applies at each age. In addition, eight Ph.D. students in developmental psychology sorted the items in the Attachment Q-set into nine piles ranging from most socially desirable to least socially desirable for each age. The sorters represented a wide range of theoretical perspectives in contemporary developmental psychology, from behaviorally oriented child clinicians to eclectic cognitive social learning theorists to psychodynamically trained personality researchers.

The mean item placements of each item in the security, dependency, and sociability, sorts for 12 and 36 months are listed in the Appendix. There was exceptionally good agreement as to the Q-sort definitions of each construct at each age. The mean correlations among criterion sorts ranged from .70 to .80. There were no obvious patterns of agreement or disagreement related to the sorter's theoretical orientation. The sorts for each construct were averaged to provide a composite definition of each construct at each age. The reliabilities of these composites were > .95 for each construct at each age. Thus the Attachment Q-set enables us to develop extremely reliable consensual definitions of these constructs from a relatively small number of sorters.

As indicated above, these criterion sorts can be used to assign scores to individual subjects by computing the correlation between the criterion sort and the array of scores assigned as a description of the subject in question. The Q-sort method clearly provides an economical means of informing and standardizing assessment of these abstract constructs across laboratories.

Conceptual similarity. Correlations across criterion sorts indicate the degree of conceptual similarity among different constructs. In the present data, security and dependency, which have often been used interchangeably and occasionally as opposites, are conceptually orthogonal at 12 months (r = - .09) and somewhat negatively correlated when they are defined with reference to 36-month-olds (r = - .36). This kind of data can inform developmental theory and also provide both justification and procedures for developing age appropriate assessment of important constructs.

Our criterion sorts also indicate that security and sociability are conceptualized similarly for both 12- and 36-month-olds (r = .82, .88). That is, psychologists are somewhat more sensitive to (or more aware of) normative trends in attachment behavior than in sociability.

Deane & Waters (1984) present analyses of item differences among these conceptual sorts as well as correlations among security, dependency, and sociability scores assigned to individual subjects. They also describe procedures for evaluating the effects of social desirability variance on conceptual sorts and for assigning scores to one criterion sort while holding variance in the other sorts constant.

In brief, detailed analysis of criterion sorts can sensitize us to our own theoretical expectations and provide a foundation for empirical research that leads to improved conceptual definitions and better measurement.

Criterion sorts were provided by Everett Waters, Alan Sroufe, Mary Ainsworth, Mary Blehar, Mary Main, Donelda Stayton, Donna Weston, and Joan Stevenson-Hinde (security: 12 months); Everett Waters, Mary Blehar, Alan Sroufe, Brian Vaughn, Inge Bretherton, Robert Marvin, Leah Matas, and Alicia Lieberman (security: 36 months); Willard Hartup, Joseph Campos, Eleanor Maccoby, Carl Corter, Susan Goldberg, Margaret Ricks, Brian Vaughn, and Donna Bradshaw (dependency: 12 months); John Masters, Tiffany Field, Jay Belsky, Mavis Hetherington, Ken Rubin, Kathleen Deane, Barbara Caparulo, and Mary Main (dependency: 36 months); Ross Parke, Allison Clarke-Stewart, George Morgan, Carol Eckerman, Mary Rothbart, Hill Goldsmith, Marguerite Stevenson, and Marilyn Svejda (sociability: 12 months); Wanda Bronson, Marian Radke-Yarrow, Thomas Berndt, Sandra Scarr, Wyndol Furman, Dante Cicchetti, Karen Rosen, and Sarah Sternglanz (sociability: 36 months); JoEllen Hoffman, Carol Friedman, Terri Lomenick, Pat Murray, Miriam Kramer, Stacy Vedder, Joyce Prigot, and an anonymous student, State University of New York (social desirability: 12 months); and Virginia Tinsley, Donna Cox, Lucille Anderson, Charlotte Jungblat, Carol Andreason, Marlene Zelek, Jan Pederson, and Paul Dores (social desirability: 36 months).

Note: Experience has shown that the Attachment Q-set does not contain enough socially desirable items from non-attachment domains to effectively assess this response style without also including some valid security and dependency variance. Accordingly, the social desirability criterion sort is no longer used and the item placements for this construct have been removed from Appendix 1 so the data for the other constructs will fit on pages in portrait format.

Evaluation

Even at this early stage in our research it is evident that the Q-sort method holds considerable promise for assessing attachment relationships in a wide range of ages and, insofar as parents continue to provide good data, in quite large samples. The method shows every indication of meeting each of the criteria outlined above for assessing attachment from an ethological/control systems perspective. The level of behavioral detail and context specification achieved in the Attachment Q-set items affords considerable structural fidelity vis a vis a control system view of the secure base phenomenon.

The Attachment Q-set and the data analytic procedures available for use with Q-sort data seem well suited to the task of examining relationships among affect, cognition, and behavior in the attachment domain. The Q-sort method also lends itself to analyses of both quantitative and qualitative developmental change. Finally, the Q-sort method is ideally suited to assessment of discriminant validity and to statistical control of sources of variance in scores computed from criterion sorts.

While data obtained via the Attachment Q-set are not merely trait ratings, they are not quite behavioral observations either. They are behavior relevant, however; and they are economical to obtain. As a result, they can play an important role in research designed to find out what the answers to important questions are going to be like. It is only when we know what the answers to our important questions are like and where they are to be found that we can design strategic observational and experimental studies that can ultimately provide decisive answers.

APPENDIX 1

CRITERION SORTS FOR ATTACHMENT CONSTRUCTS

	Security		Dependency		Sociability	
ITEM TITLES'	12	36	12	36	12	36
1. Remains fearful of moving toys or animals .	4.00	3.38	5.87	6.12	3.37	3.25
2. Eager to demonstrate songs, games, or other behavior	5.37	6.62	3.75	4.37	7.75	6.50
3. Predominant mood is happy.	8.25	7.87	3.62	3.25	7.25	8.62
4. Easily comforted by adult	8.25	8.12	3.37	3.75	6.87	7.00
5. Approaches adult to interact.	4.50	4.12	7.25	7.00	6.62	5.00
6. Prefers tasks and activities that are not difficult	4.50	4.37	5.75	6.62	4.87	4.00
7. Is often unaware of changes in adult's location or activities	3.50	4.75	1.37	2.25	3.62	6.00
8. Laughs easily with observer.	6.25	5.87	3.00	3.37	8.00	7.62
9. Does not babble or talk when playing alone .	4.00	4.50	5.12	5.12	3.12	4.25
10. Avoids or rejects new people. .	3.62	3.37	6.50	7.62	1.37	2.75
11. Does not recognize distress in adult	3.75	3.37	4.00	4.62	3.25	3.62
12. Bouts of exploration and play away from the adult are brief	2.62	1.75	7.87	7.50	5.00	3.87
13. Becomes bored quickly .	4.12	4.25	5.50	5.62	4.00	2.62
14. Does not accept adult's affectionate behaviors to others	3.87	3.87	6.25	7.25	3.87	2.87

15. Prefers female adults.	5.12	5.12	6.00	5.37	4.75	4.75
16. Is upset by negative evaluations or 6.37 disapproval from adult	5.50	7.75	7.25	5.75	5.25	
17. Does not share willingly .	4.12	3.25	5.50	5.62	2.25	2.62
18. Actively solicits comforting from adult when distressed5	8.37	8.00	6.50	7.75	7.37	6.2
19. Explores objects thoroughly	7.00	6.25	4.25	3.75	4.75	6.62
20. Distressed by separation at home .	2.62	1.37	7.75	7.62	4.12	3.50
21. Is indifferent to observer's invitations to play	3.75	3.12	4.75	6.25	1.37	3.50
22. Easily distracted from distress. .	6.25	6.50	3.37	3.00	6.37	7.12
23. Has good endurance, is not easily tired.	5.75	5.75	4.00	4.37	5.37	6.50
24.Proximity/exploration/proximity cycles are evident in 1/2-1-hour observations .	8.00	5.62	5.62	4.25	6.50	5.25
25. Is affectively responsive and expressive.	7.87	7.87	4.87	3.87	8.25	7.87
26. Does not cry hard from minor injuries.	6.12	6.25	2.25	2.25	5.12	7.00
27. Is careful with toys	6.00	6.25	4.62	4.87	4.87	6.50
28. Is not adaptable when moved from one activity to another	3.25	4.13	6.37	6.62	3.87	2.75
29. Cries to prevent separation	4.00	3.25	8.25	7.75	4.25	3.62
30. Is responsive to distress in adult when wary.	6.25	6.87	4.62	4.12	7.25	7.12
31. Does not look to adult for reassurance .	1.75	2.37	1.75	2.62	2.62	3.75
32. Initiates interaction with familiarized adults	6.87	7.12	3.25	3.12	8.12	7.12
33. Maximum good mood requires adult's presence	5.00	4.00	7.25	7.37	5.75	3.12
34. Does not attempt to approach or follow when adult moves away	3.87	4.12	2.37	1.75	3.12	5.25
35. Prefers to be comforted by adult .	7.75	6.62	7.62	7.00	4.62	4.62
36. Greets adult spontaneously .	8.75	8.37	4.87	4.00	8.75	7.25
37. Is demanding and impatient.	3.00	2.75	7.87	7.62	4.25	1.87
38. Aware of social environment.	6.87	7.37	6.00	5.87	8.12	7.50
39. Hesitates or does not repeat previously prohibited behavior	6.50	6.75	5.50	5.25	5.62	6.25
40. Acts to maintain social interaction	7.25	6.75	4.37	3.75	8.62	8.00
41. Is flexible in trying to communicate clearly with adults	7.12	7.62	3.12	4.12	7.00	7.00
42. Is independent with most adults .	5.75	6.87	2.00	1.12	5.87	7.37
43. Returns from exploration and play often spontaneous at home	7.75	6.87	7.00	6.12	7.25	5.62
44. Does not solicit or enjoy physical contact with nonfamily adults	4.62	4.25	5.00	5.87	2.50	3.37

45. Actively solicits assistance or comfort after minor injury	7.37	6.37	7.12	7.75	6.62	5.25
46. Gross motor control is not smooth and coordinated	4.50	4.87	5.25	5.25	4.87	4.00
47. Interacts directly with adults. .	7.25	7.75	5.50	5.37	8.50	6.87
48. Lacks self-confidence	3.50	2.12	7.50	7.62	4.00	2.37
49. Prefers realistic play. .	5.00	4.37	4.62	5.00	4.37	4.25
50. Behaves in a nurturant or parental way toward toys during play	6.12	7.00	4.37	6.37	6.37	6.37
51. Doesn’t accept adult's assurances when wary in familiar contexts	2.25	2.62	6.25	5.87	2.25	4.12
52. Transition from exploration to proximity and contact is not executed smoothly	2.00	2.12	5.62	5.12	3.00	3.00
53. Does not solicit or enjoy affectionate physical contact with adult	1.50	2.12	2.50	3.25	1.62	1.75
54. Expects adult will be unresponsive	1.37	1.25	5.62	5.37	2.62	2.00
55. Cries in response to separation	5.62	3.87	8.37	7.87	4.87	3.75
56. Does not display tension movements	6.25	5.25	3.62	3.12	5.50	5.87
57. Average activity level is high. .	5.12	5.12	4.37	4.12	4.75	5.75
58. Is not compliant with adult control	2.75	2.75	4.37	4.37	2.75	2.62
59. Is attracted to novelty rather than familiarity	6.62	6.37	3.37	2.62	6.12	7.50
60. Sleeps on a regular schedule. .	5.75	5.25	4.75	4.50	4.75	6.12
61. Is not bolder or more confident to play when adult is nearby	3.12	3.75	2.37	3.50	4.00	4.50
62. Becomes distressed when social interaction is blocked or becomes difficult	3.87	3.87	7.00	7.50	5.62	3.37
63. Becomes distressed when adult moves away .	3.87	3.25	8.25	8.00	4.75	3.62
64. Does not solicit or enjoy playful physical contact with adult	2.25	1.87	3.12	4.12	1.75	2.25
65. Object oriented in play preferences	3.12	3.87	3.62	4.00	2.00	3.37
66. Does not persist when nonsocial goals are blocked. .	3.50	3.62	5.87	6.50	4.62	3.37
67. Sleeps lightly (even if regularly)	4.75	4.75	5.25	5.13	4.87	4.50
68. Transition from proximity and contact to exploration is not executed smoothly	2.00	2.25	7.00	6.75	3.25	3.62
69. Is independent with adult	5.12	7.12	1.12	1.12	5.87	7.75
70. Is indirect or hesitant in making observations or requests	3.75	3.37	5.12	5.00	3.25	3.50
71. Prefers animate toys	5.25	5.50	4.62	4.75	6.00	5.25
72. Does not stay closer to adult in . . unfamiliar settings	2.50	4.37	1.75	1.25	4.12	5.37
73. Accepts assistance	7.37	7.00	6.12	6.37	6.87	7.75

75. Cries often (regardless of intensity or duration)	2.50	2.87	7.12	7.62	3.50	3.12
76. Expresses enjoyment of accomplishing or achieving	7.00	7.12	4.00	3.87	6.87	7.87
77. Affective sharing occurs during play	8.25	8.12	4.37	4.62	8.62	7.37
78. Does not restart crying readily after crying and calming down	7.75	7.00	2.50	3.00	5.50	6.75
79. Imitates observer's behavior	5.87	5.62	3.75	3.75	7.50	5.75
80. Is more tolerant of self-initiated separation than of adult-initiated separation	5.75	6.25	6.62	6.25	5.37	5.37
81. Is creative with objects or social roles in play	5.62	6.25	3.87	4.12	5.00	7.87
82. Easily becomes angry with adult	2.00	3.37	5.62	6.87	3.25	2.00
83. Recovers from minor injuries slowly .	4.75	4.50	6.00	6.12	4.62	3.62
84. Does not adapt active play to avoid hurting adult. .	3.75	2.87	5.12	4.75	3.87	2.00
85. Requires encouragement to keep constructively occupied	3.25	3.75	7.37	7.12	4.87	3.00
86. Does not accept adult's assurances when wary in unfamiliar places	2.00	2.75	6.50	6.75	2.00	3.25
87. Does not laugh easily with adult	3.00	3.50	5.37	5.37	1.62	3.12
88. Imitates adult's behavior	5.87	6.62	5.50	4.25	7.62	6.25
89. Proximity/exploration/proximity cycles are evident in 3-5-hour observations	6.87	8.25	4.87	4.62	6.12	6.12
90. Shows signs of self-control .	6.37	7.87	3.50	4.25	5.62	7.12
91. Rarely asks for help. .	3.87	4.37	1.37	1.75	3.12	5.00
92. Does not become angry with toys .	6.62	6.12	4.25	3.87	5.00	6.75
93. Accepts adult's attention to others	6.75	6.75	3.50	2.25	6.37	6.87
94. Returns from exploration and play are often spontaneous in unfamiliar environments	7.75	7.00	6.50	6.50	6.75	5.75
95. Child's observations and requests are often difficult to understand	3.87	4.12	4.87	5.25	3.87	3.50
96. Is obedient when adults give instructions.	7.37	7.00	5.50	5.12	5.87	6.00
97. Is not wary of new objects .	5.37	5.75	2.75	2.25	6.12	6.00
98. Does not prefer physical contact with adult .	2.00	3.00	2.50	3.50	3.37	4.37
99. Fine motor manipulation is not skillful.	4.87	4.87	5.25	5.12	4.87	4.87
100. Does not combine several objects in play.	4.50	4.50	5.00	4.75	4.75	4.62

NOTE.-The criterion scores were constructed by averaging the scores assigned to the items by each of the sorters. The items were sorted into nine piles with 5, 8, 12, 16, 18, 12, 16, 8, and 5 items, respectively. Items in pile 1 (least characteristic) received scores of 1, and so forth. a The complete items and sorting instructions are available on-line at www.johnbowlby.com.