

9/88

MEASURING TEMPERAMENT IN A LARGE CROSS SECTIONAL SURVEY:
RELIABILITY AND VALIDITY FOR CHILDREN OF THE NLS YOUTH*

By

Elizabeth G. Menaghan and Toby L. Parcel

Department of Sociology

The Ohio State University

*Funding for this research was provided by the Center for Human Resource Research at The Ohio State University.

MEASURING TEMPERAMENT IN A LARGE CROSS SECTIONAL SURVEY:
RELIABILITY AND VALIDITY FOR CHILDREN OF THE NLS YOUTH

Many developmental psychologists traditionally have studied the patterning of child behavior with small samples of children, relying on replication of findings across studies to provide evidence for external validity. Such small samples have inherent limitations: rather than estimating parameters to assess the strength of specific relationships, investigators must rely more heavily on statistical tests pointing to the existence of differences between or among groups of children, or the existence of association between two variables, possibly controlling for a few covariates. In addition, small samples cannot be appropriately divided to portray the dynamics of statistical interaction, even when theory suggests that understanding the magnitude of such interaction may be critical to understanding the issues. While some recent efforts have been devoted to studying developmental processes with larger samples (Ruopp and Travers, 1982), until recently such data sets were rarely available and utilized.

By the autumn of 1987, a new resource with which to study child development had become available. This resource is a survey of the children of mothers from the National Longitudinal Survey's Youth Cohort, where the survey of the children contains a number of age appropriate measures of cognitive and social development. While some child outcome measures were included in the survey in their entirety, time restrictions in a cross-sectional survey format precluded that possibility for other measures. For some, subsets of items from validated measures were used; for others, items were combined from several measures to ensure coverage of important dimensions. This study investigates the measurement properties of the set of items combined from several sources to assess the

construct of temperament. After describing the nature of the Youth Cohort itself and origins of the survey of children, we describe the scaling procedures used to construct measures from the items, report these measures' reliabilities, and describe initial evidence suggesting validity. We conclude with discussion regarding the advantages of using these measures for study of certain research questions central to the field of child development, as well as limitations inherent in the resource.

The Youth Cohort and Children of the NLSY

The Youth Cohort is a panel study of a national sample of youth who were 14-21 years old in 1979. The sample was derived from two sampling frames. The first was a cross section of youth in the population at that time and the second oversampled black, Hispanic youth, and economically disadvantaged non-black, non-Hispanic youths so as to provide additional cases for analyses of the economically disadvantaged. The cohort was followed up with interviews each year; some items were asked every year while others were asked less frequently. By the middle 1980s a wealth of information had been collected; the general constructs covered included family background information, attitudes and social psychological constructs, detailed information regarding schooling, including data derived from school records, a wealth of labor force participation information, data on alcohol and drug usage, and detailed information on marital and fertility histories. The NLSY has had a 91.8% retention rate over the duration of the panel (NLS Handbook, 1987).

Of the 4,918 young women in the sample, nearly 60 percent (n= 2,918) had become mothers by 1986. With funding from the National Institute of Child Health and Human Development (NICHD), in 1986 the Center for Human Resource Research at The Ohio State University (CHRR) was able to survey the children of the mothers from the Youth Cohort; interviews and assessments were completed for 4,971 children. Given the age of the cohort in 1979, the sampling frames from which it was derived, and the year of the child survey, these children

represent early and on time births from a cohort of disproportionately lower socioeconomic status mothers: over a third of the children assessed were born when their mothers were in their teens, and about half were born when their mothers were aged 20 through 24.

A national panel of child development experts was convened by NICHD in order to choose the measures to be included in the survey. Table 1 lists these measures, indicates whether the entire measure or a subset of items was included in the survey, and indicates the ages of the children surveyed for each measure. In all cases interviews were conducted face to face with trained field interviewers under the direction of the National Opinion Research Center (NORC) affiliated with the University of Chicago. In many cases the interviewers were the same people who had interviewed the mothers for several years prior to 1986, and thus had developed rapport with the mother and potentially with the children. The data set is publicly available from the CHRR. Of course, the possibility of linking a detailed data set of maternal characteristics with good measures of child outcomes is an exciting one for researchers from several disciplines.

In order for such potential to be realized, however, we must be assured that the measurement of the child characteristics themselves is sound. In the cases where measures have been included in the survey in their entirety, individual researchers can assess reliability of the measures for the samples they select in particular investigations, and provide evidence regarding validity within the context of the research problems they study. National norms are available with which to compare sample results, so that if biases in the study samples exist, their magnitude can be estimated and findings interpreted accordingly. In the case of the measures where subsets of items have been selected, however, these procedures are insufficient. Whether these subsets form one or more scales must be investigated; new measures must have acceptable levels of reliability and demonstrated validity. In this paper, we

develop measures of children's temperament and provide preliminary evidence regarding their validity.

MEASURING CHILDREN'S TEMPERAMENT

Temperament has been viewed as a set of relatively consistent, basic dispositions inherent in a person that underlie and modulate the expression of activity, reactivity to stimuli, emotionality (positive/negative mood), and sociability/shyness (Goldsmith et al. 1987). While such dispositions may be biologically based, the expression of those dispositions in observable behavior is influenced by the individual's experiences. There is disagreement regarding the extent to which early individual behavior differences reflect heritable and physiologically based tendencies or the effects of variations in experienced social environments (Bates 1980); as babies grow older, it is generally conceded that the expression of temperament is increasingly shaped by prior history. Nevertheless, even among adults, some researchers argue that constitutional differences in dispositions endure. Kagan (1984) views both fearful infants and introverted adults as similar in their initial behavioral and cardiovascular reactions to unfamiliar or risky situations: they inhibit behavior and show more cardiovascular arousal. Emotional tone is also considered relatively persistent across the life course, with child development specialists discussing it in terms of irritability and fussiness and mental health researchers discussing negative affectivity (Costa, McCrae and Arenberg, 1980).

To tap dimensions of temperament, the 1986 Child Assessment utilized mothers' reports of children's behavioral styles. This approach is one of two major approaches to the measurement of temperament; the other utilizes assessments by trained raters based on direct observation of infants and children at home or in a laboratory. Each of these approaches has advantages and limitations: assessments by trained raters generally sample a very small time period of a child's life and may be affected by transitory conditions; in contrast, those of parents (or occasionally, teachers) can be based on long

familiarity with the child. On the other hand, observations by trained raters are viewed as more objective, while mothers' ratings in particular have been viewed as subjective reflections of maternal attitudes, characteristics, and perceptions as much as, if not more than, reflections of child characteristics.

In support of this view of maternal ratings, research has shown that mothers' ratings of their babies' temperaments are predictable from maternal characteristics and attitudes assessed prior to the child's birth (Vaughn, Bradley, Joffe, Seifer, and Barglow, 1987). Other evidence suggests, however, that correlations between maternal characteristics and maternally-reported temperament are not due simply to biased maternal perceptions: Matheny, Wilson, and Thoben (1987) report that observation-based lab assessments are also predictable from maternal characteristics. They also note relatively high agreement between the maternal reports and the ratings that observers made after observing children during standardized laboratory-based assessments (for example, the correlation was .52 for 24 month olds).

Thus, it is incorrect to view maternal ratings primarily as reflections of mothers' rather than children's characteristics. Nevertheless, maternal ratings do seem to be affected by maternal characteristics and situation: multivariate analysis suggested that with laboratory-rated temperament held constant, maternal ratings were related to maternal characteristics and household conditions. Such relationships between maternal or household characteristics and maternal temperament ratings may suggest that mothers' characteristics and household conditions affect mother-child interaction and produce changes in how the child behaves at home with mother as compared with how he or she reacts to a laboratory situation.

In summary, maternal reports of children's temperaments are associated with other, standardized judgments of child characteristics, as well as with maternal and household conditions that shape mother-child interaction. These findings suggest that the strategy of asking for maternal reports of items

relating to child temperament may be a valid one. Our analysis suggests the extent to which the findings derived from the procedure are useful.

For the NLSY assessment, age-appropriate items were adapted from published sources to reflect major dimensions of temperament and be readily understandable even for respondents with low educational attainment; mothers read the items and indicated the appropriate response in a self-administered questionnaire. Because the items asked varied by child age, we shall discuss scale construction separately for three age groups: babies under a year of age; babies and toddlers under age two; and older children age two through six years old. We then evaluate the validity of these scales by examining their associations with other child measures and with household and maternal characteristics.

Two additional sets of items on the child assessment can be used to assess validity of the temperament measures we describe: measures of mental ability and interviewer assessments of the quality of the interviewer-child interaction during testing. The interviewers conducted age-varied cognitive assessments with all children at least eight months old. The interviewer assessment was taken at the end of the interview, and in contrast to the maternal ratings of the child's usual way of acting, measured how the child interacted with a relative stranger and reacted to a novel set of demands. Since this information is independent of maternal reports, it can be used to assess the validity of the maternal reports. We therefore also construct measures from these interviewer ratings.

CONSTRUCTION OF TEMPERAMENT MEASURES

We have noted that the NLSY mothers are a disproportionately lower socioeconomic status group. To correct for this, all analyses were conducted with weighted samples, where the weights were constructed to reflect a nationally representative sample of households. The fact that the sample contains a disproportionate number of children from lower socioeconomic status households therefore does not bias the analyses since these cases have been weighted down

to reflect their relative frequency in the larger population. Their presence in the sample does help to increase the reliability of the estimates derived using the data, particularly if one is interested in estimates for such subgroups.

Younger Children. As indicated above, the items asked of mothers varied depending on the child's age. Mothers of children under two years of age were asked about their child's positive affect, fearfulness to novel stimuli, and general irritability or fussiness; for babies under a year, they were also asked about the child's overall activity level and the rhythmicity/predictability of their babies' routines. Items were adapted from Rothbart's Infant Behavior Questionnaire (Rothbart and Derryberry, 1984); Bates' Infant Characteristics Questionnaire (Bates et al., 1979), and Carey and McDevitt's Infant Temperament Questionnaire (Carey and McDevitt, 1977).

Table 2 reports the means and standard deviations of items asked of these two groups of children, and the results of two factor solutions. The first solution is for the six items asked only for infants 0 to 11 months, intended to assess overall activity/quietness and rhythmicity. Responses to these questions were in terms of the frequency each activity is observed where 1 represented almost never, 2 represented less than half the time, 3 represented half the time, 4 represented more than half the time and 5 represented almost always. The analysis is based on the cases that had valid responses on all six items; the unweighted number of cases was 560. Variations in means across items suggest that mothers are more likely to perceive consistency in their infants' daily moods, sleep and hunger patterns than they are to perceive high levels of activity.

We subjected these six items to a factor analysis using principal axis factoring and an oblique rotation. As expected, two factors met the eigenvalue criterion and both were substantively interpretable. We interpreted an item as loading on a factor if it loaded at an absolute level of .35 or greater. The first factor we named Active; both squirms/kicks and

waves arms during feeding as well as moves in crib during sleep all loaded on this factor. The second factor we named Predictable since the additional three items, each referring to consistency of mood or physical needs, all loaded on this factor. Although the oblique rotation permits factors to correlate, the two factors are essentially orthogonal; the correlation among the underlying factors is .03. We formed factor-based scales corresponding to this factor solution, where items were transformed into standard scores prior to summing items to create the scales. Alpha reliabilities for the composites are Active: .66 and Predictable: .64. These relatively low levels of reliability are produced by the moderate inter-item correlations and the small number of items available for these scales.

The second factor analysis reported in Table 2 is for the 11 items asked of mothers with children aged 0 through 23 months, intended to tap positive/negative affect, fearfulness, and irritability or fussiness. These items were coded in terms of the frequency each activity is observed as indicated above, except for the last two items. Responses to frequency of fussy times include 1, almost never; 2, once or twice a day; 3, couple of times in AM and PM; 4, several times a day; 5, almost every hour. General quantity of fussiness is coded 1, almost never; 2, less than average; 3, about average; 4, more than average; 5, almost always. The analysis is based on the 1,139 cases of infants and toddlers between 0 and 23 months who had valid responses on the complete set of 11 items. Inspection of the means shows that mothers were unlikely to report high levels of fearful crying or fussiness, and tended to regard their children as smilers, especially when they played with them.

Using the same factor analytic procedures as described above, the initial solution suggested that three factors met the eigenvalue criterion. However, the third factor contained only two items: frequency of fussy times and general quantity of fussiness. Since these items used a different response format from the remaining items, we attributed the three factor solution to varying question

format, and constrained a second factor solution to two factors. In this second, constrained solution, we named the first factor Fearful-Fussy because loading on it are items suggesting fear in reaction to strangers, animals, maternal absence, the doctor, and loud sounds, as well as the items related to fussiness. The second factor we named Smiley because all three of the items relating to smiling loaded on it. Although the oblique rotation permits correlation among factors, Smiley and Fearful were essentially orthogonal, with the correlation among the underlying factors .02. The factor-based composites derived from this analysis had reliabilities of .67 (Fearful-Fussy) and .75 (Smiley). For the babies under age one with valid scores on all four infant-toddler scales, more fearful babies were viewed as less predictable ($r = -.15$, $p < .001$) and more active ($r = .24$, $p < .001$), while smiley babies were both more predictable ($r = .30$, $p < .001$) and more active ($r = .18$, $p < .001$).

Older Children. For children aged two through six years eleven months, mothers were asked about a range of child behaviors, including their compliance with/resistance to parental requests; shyness with new peers or adults; positive/negative affect; quality of play with other children; and quality of mother-child interaction. The items intended to tap mother-child interaction were adapted from Everett Water's Attachment Q-Sort (Waters and Deane, 1985). Responses to the 19 maternal report items are in terms of the frequency with which the behavior is observed. One item tapping whether the child sleeps through the night was omitted from the analysis because of very low item variance. The final two items included in the factor analysis are the interviewers' assessments of how shy/sociable the children were at the beginning and at the end of the child assessments. Responses to these items were coded on a one to five scale where one represented not at all shy and anxious, or being sociable and friendly, and five represented being extremely shy/quiet/withdrawn. While these last two items are not maternal ratings, they do represent ratings by trained observers of an important temperament

dimension, shyness/sociability. We hypothesized that they might contribute to the construction of temperament relevant scales to be used in this data set and included them in the factor analysis.

Table 3 presents the item statistics and factor analytic results for these 21 items. Several items have relatively little variation in this age group, with most children reported to eat what is served, and smile or laugh easily. Variation in the means associated with the items suggests that high proportions of mothers perceive that their children eat what is served, share toys with other children, and try to help when the mother appears upset, while fewer find it hard to soothe the child or perceive the child to be upset if they are left alone. Initial shyness with the interviewer is relatively low, and on average the children had become more sociable by the end of the assessment.

Factor analyses were based on the cases with valid responses for all 21 items; the unweighted number of cases was 2,433. Reflecting the broad coverage and diverse content of the item pool, inter-item correlations were relatively low, and the estimated communality of items ranged from .11 to .44, with an average estimated communality of .25. The initial factor solution yielded seven factors meeting the eigenvalue criterion of 1; three had eigenvalues above 1.5, while four more were below 1.5. Extraction and rotation of all seven factors produced extremely narrow-band factors, each identified by only a few items; only three factors had more than two items loading above .35. These findings suggested that the variation in the data could be well summarized by three factors, a decision that might also facilitate building reliable scales. A second factor solution was constrained to produce three factors; these results are displayed in Table 3.

The first factor we named Compliant because loading on it are items referring to eating what is served and turning off the TV when asked to do so, either immediately or after protest, going to bed after protest, and sharing toys with other children. This factor is conceptually related to the

adaptability/rhythmicity construct from which a number of temperament items were derived, and which is reflected in the scale Predictable for infants 0-11 months. For older children, Compliant may also reflect the degree of parent-child conflict around scheduling. A related item, asking about how readily the child went to bed when first told to do so, did not load on this factor (factor-item correlation was $-.18$).

The second factor we named Shy; loading on it are shyness in meeting new children and adults and the two interviewer assessments of degree of shyness at the start and conclusion of the interviews. This shyness/sociability factor is conceptually related to the items tapping fearfulness to new people and situations that were asked for younger children.

The third factor we named Dependent-Demanding; loading on it are items referring to the child's interaction with mother (being demanding or impatient even when the mother is busy, becoming upset when the mother leaves the room, staying close to the mother and/or keeping her in sight while playing, wanting the mother to help with child's activities, and the mother having difficulty soothing the child), peer interaction (aggressing against other children), and crying at minor discomfort. There is an anxious, disorganized undertone to this factor that links it most closely to the temperamental construct of negative affect; it may also partially reflect an insecure parent-child attachment in which the child is perceived as seeking closeness to mother in a way that may irritate her. Interestingly, more attractive attachment-related behaviors had lower or near-zero loadings on this factor; liking to imitate mother: $.33$, and trying to comfort her when she is upset: $.01$.

In all, four items in the item pool did not load above $.35$ on any of these three factors. These were scattered across several conceptual domains: predictability/rhythmicity (going to bed without protest), positive affect (smiling easily), and positive mother-child interaction (child likes to imitate mother, tries to help when she is upset). The pool of items were

insufficient to yield separate multiple-item scales for these constructs.

Alpha reliabilities for the factor based scales associated with the three factors are: Compliant: .64; Shy: .68; Dependent-Demanding: .63. These relatively low measures of internal consistency reflect the low to moderate intercorrelations among the items. Compliant, Shy, and Dependent-Demanding were only moderately inter-correlated, with correlations among the underlying factors all below .20. Similarly, correlations among the constructed scales were all below .30: children viewed as more dependent-demanding were also seen as more shy with new people ($r = .18$, $p < .001$) and less compliant at home ($r = -.27$, $p < .001$); shy children were also somewhat less compliant ($r = -.09$, $p < .001$).

VALIDITY

In order to provide evidence on construct validity, we assess relationships between each scale and a number of dimensions we would expect on the basis of previous theory and research to be related to them. Theory predicts that child temperament will be related to reactions to a novel situation such as the testing situation with the interviewer. Children who are fearful, irritable, shy, or dependent-demanding would be expected to receive lower interviewer ratings than children who are readily compliant and smiley. To the extent that maternal ratings of their children's temperament capture real differences in behavioral style, we should find that maternal ratings of child temperament and interviewer assessments regarding the child's handling of the interview situation are associated.

Studies have also shown some measures of temperament to be related to cognitive development (Bates 1980; Campos 1983). While differences in overall activity or rhythmicity in the first year may have no strong or necessary impacts on later development, affective tone and fearfulness or shyness are expected to shape interaction and development. A baby or child who smiles readily and responds positively to an adult's or child's interaction efforts helps to extend that involvement and stimulation, while the responses of an

irritable easily frustrated, or fearful child may short-circuit the interaction or constrain the child's ability to learn from the interaction. Consistent with this argument, Campos (1983) summarizes research evidence that positive affect is positively related to language development and cognitive performance, while temperamental "difficulty" is associated with slower cognitive development. Therefore, we expect that child temperament will be related to assessed verbal and cognitive development. Since such effects may take time to emerge, these associations should be stronger at preschool and later ages.

There is also evidence linking temperamental measures to later behavior problems. Cameron (1978) found that a temperament composite of high persistence, low adaptability, negative mood, and withdrawal from new experiences, in combination with an environmental composite of parental conflict about childrearing, inconsistent discipline, and mother rejection predicted later behavior problems. Korn (1984) similarly found that difficult temperament, adverse parental attitudes and practices, and an accumulation of stressful events were risk factors that in combination tended to produce behavior disorders. Finally, we expect that an earlier self report of maternal adult shyness will correlate with perceptions that the child is shy because evidence suggests that this dimension of temperament is highly heritable (Kagan, Reznick, Clarke, Snidman, and Garcia-Coll, 1984), and because mothers may perceive their children to have similar personality traits to themselves, independent of temperamental propensity.

Measuring Interviewer Ratings. We derive measures of interviewer assessments in order to evaluate the validity of the temperament scales discussed above. Table 4 presents two factor analyses based on the five additional interviewer assessments (other than ratings of shyness incorporated into a previous solution) for the children eight months and older. The items tap attitude toward testing, rapport with the interviewer, perseverance/persistence, cooperation, and motivation; all are scored on a one to five scale

where one indicated poor, three indicated average and five indicated excellent. Consistent with the analyses presented thus far, factor analyses were conducted separately by age group; there were 790 children between 8 and 23 months of age and 2,609 children between 24 and 83 months who had valid responses on the interviewer items. Inspection of the means suggests that the older children get noticeably higher interviewer ratings on these items than the younger children, possibly because behaviors suggesting cooperation, motivation, persistence, etc., are more frequent among older than younger children. Younger children are perceived to have greater rapport with the interviewer than to be cooperative and motivated, while there is minimal variation among the means of the items for the older children. In each case, however, the five items are highly intercorrelated and all load on a single factor. We have named these scales Positive Impression. Alpha reliabilities for the scales are .94 (ages 8-23 months) and .95 (ages 2 to 6 years, 11 months).

Evidence for Construct Validity

Since previous investigations of temperament have used age-homogeneous samples in assessing validity we also examine age patterns in temperament scales. Age (in months) is significantly related to scores on every temperament measure. Among the babies and toddlers, older children are perceived as somewhat more active ($r=.18$), more predictable ($r=.17$), and more fearful-fussy ($r=.18$) than younger children. Age is strongly linked to descriptions of smileyness ($r=.43$); additional analysis showed that this strong relationship is partly due to the very low and homogeneous scores of the youngest infants (under 3 months), for whom frequent smiling is developmentally less expectable. Among the children between two and six years, older children were seen as more compliant ($r=.22$), less shy ($r= -.15$), and less dependent-demanding ($r = -.20$). Because we wished to assess hypotheses bearing on validity without contamination from associations with age, we control for age when reporting correlations with other constructs, and we exclude infants under 3 months when

assessing relationships with ratings of Smiley.

Table 5 displays the partial correlations among the temperament scales and the variables expected to relate to them. Relationships between maternal descriptions and interviewers' impressions are moderate. Among infants eight months or older, those who make a positive impression on the interviewer are perceived by their mothers as less active, more predictable, less fearful-fussy, and more smiley. Among children two through six years, those perceived by their mothers as more readily compliant, less shy, and less dependent-demanding are more likely to make a positive impression on the interviewer.

The relationships between temperament and cognitive and verbal skills are also generally consistent with expectations. Kagan's Memory for Location test, which was used to assess short-term memory for babies 8 through 47 months, is not significantly related to infant activity, predictability, or smileyness, but is inversely related to fearfulness. For older children, cognitive assessments varied with the child's age (see Table 1). Two year olds took the Kagan test, three year olds took both the Kagan test and the Peabody Picture Vocabulary Test (PPVT), and the four to six year-olds took only the PPVT. Kagan's test had high means and low variation among the older children, and it was not related to the child temperament measures when child age was statistically controlled. However, the expectation that temperament would affect interaction and stimulation, and therefore be associated with more advanced language development, was supported for the older children: those higher on Compliant and lower on Shy and Dependent-Demanding demonstrated greater vocabulary knowledge.

In addition to being associated with child temperament as just described, the interviewer ratings were also associated with children's performance on both the short-term memory and vocabulary assessments: those with whom the interviewer was able to establish rapport and elicit cooperation score higher on the assessments. Thus, the observed relationships between more difficult

temperament and lower cognitive and verbal scores may partially reflect a downward bias in the assessment of children's capacities when they had more difficulty or discomfort in interacting with the interviewer. To assess the extent of this bias, we estimated the partial correlations between temperament and cognitive and verbal measures controlling for the interviewers' assessments as well as child age. With these controls, fearfulness remains inversely related to memory for location ($r = -.14$). Performance on the Memory for Location test remained nonsignificantly related to compliance (partial $r = .02$) and dependence-demandingness (partial $r = .02$) for these older children; but shyness was associated with somewhat better performance (partial $r = .08$, $p < .01$). Controlling for interviewer impressions, performance on the PPVT was still significantly associated with child compliance (partial $r = .16$) and child dependence-demandingness (partial $r = -.21$), but the inverse relationship with shyness was reduced to insignificance (partial $r = -.00$). Maternal shyness was unrelated to early infant ratings of activity and predictability, or to child ratings of Compliant or Dependent-Demanding; but infants of shy mothers were described as somewhat more fearful-fussy and less smiley, and children of shy mothers were described as more shy themselves.

Table 5 also displays relationships between temperament and current behavior problems. Since behavior problems were not assessed for children under age 4, this relationship can only be examined for the older children's scales. Mothers of children four years old and older had responded to an abridged version of Achenbach's Child Behavior Checklist; this abridged scale has been shown to have high reliability and reasonable evidence for validity (Parcel and Menaghan, 1988). The older children's reported behavior problems were strongly associated with mothers' views of them as dependent-demanding and noncompliant, and more weakly related to reported shyness.

Taken together, these findings provide evidence that the constructed temperament scales behave in accord with theoretical expectations, and reflect

aspects of child temperament and its expression in early interaction. Besides reflecting inherent differences in child temperament, however, these maternal ratings of temperament may also be affected by characteristics of the mother and the household environment that mother and child share. Our assessment of validity therefore includes examination of these relationships.

Relationships between Maternal Situation and Maternal Ratings

Tables 6 and 7 display the partial correlations between temperament variables and background characteristics of mother, child, and household, controlling for child age. We expect children from homes of higher socioeconomic status to have more favorable temperament ratings than children from homes of lower socioeconomic status. These latter homes will have lower levels of material comforts than advantaged households, with attendant difficulty in meeting basic needs of physical survival. Thus, few additional resources are available to nurture children in ways conducive to the expression of positive temperament. Children from more socioeconomically advantaged homes are likely to have access to greater parental resources, both material and psychological, since these resources enable parents to establish a more economically secure household with less strain. This "surplus" of resources may enable parents to establish positive parent-child relationships that contribute to the development of more benign expression of temperamental characteristics.

Infant and Toddler Temperament. With child age controlled, young babies' overall activity level and predictability had weak but consistent associations with family economic and occupational conditions and background characteristics (Table 6). The babies perceived as more active come from households with lower income levels, where mothers have lower levels of educational attainment and lower levels of cognitive achievement. Babies described as less predictable come from households with lower income levels, households more likely to be below the poverty threshold, and households with unmarried mothers and mothers who have lower levels of cognitive achievement.

Non-white mothers perceive their babies as more active and less predictable.

Similarly, mothers who describe their children as fearful-fussy have lower household incomes, are more likely to be in poverty, less likely to be married or employed, and if employed are more likely to have a lower occupational status; they are more likely to be minority, lower in educational attainment and cognitive achievement, and younger at the birth of the child. In contrast, smileyness is not statistically associated with any maternal variable with child age controlled. Regarding child characteristics, we find that boys are somewhat less likely to be described as smilers. Later born children are described as more active, more fearful-fussy, and less smiley than first born children. We would generally expect that later born children would appear more sociable than first born children owing to the more complex social environment into which they were born, thus prompting the early development of social skills. In this sample later born children necessarily have mothers who were young at the births of their first children, and thus are less likely to have the educational and financial resources conducive to the positive development of temperament when compared with the mothers of the first born children, many of whom are older, more educated and more financially secure.

Child Temperament. Among older children, there are similar weak but consistent linkages between maternal ratings and household and maternal resources. Maternal perceptions that the child is compliant are associated with higher household income, lower likelihood of being in poverty, greater chance of being married, higher pay and occupational status if employed, higher educational attainment and cognitive achievement and lower chance of being non-white. Relationships are less prevalent with Shy. Children perceived as shy are somewhat more likely to have married mothers, who if they work are in lower status jobs, and to have mothers who have low educational attainment.

Children perceived as dependent-demanding are more likely to come from socioeconomically disadvantaged households in terms of total household income

and poverty status. Their mothers are more likely to be non-white, less likely to be married, less likely to be employed, and if employed more likely to be employed at lower rates of pay and in lower status jobs; these mothers have both lower levels of educational attainment and of cognitive achievement. Turning to child characteristics, boys are viewed as less compliant and later borns are viewed as more shy and dependent-demanding. While we would ordinarily expect later born children to be less dependent-demanding, as noted above, these later born children are likely members of households less able to provide the social context to promote positive expressions of temperament.

INTERPRETATIONS AND CONCLUSIONS

Evidence regarding the validity of the temperament scales is uneven. The scales for infants tapping activity and predictability do not show strong construct validity. They are sparsely related to household and maternal characteristics, where the directions of the significant relationships suggest that mothers from socioeconomically disadvantaged households perceive their babies to be more active and less predictable than mothers from more advantaged households.

For two scales--babies' and infants' smileyness and older children's shyness-- there is good evidence for construct validity and relatively weak linkages with household and maternal characteristics. With child age controlled, children described as smiley are somewhat more likely to be girls, have mothers who are less shy, and make better impressions on the interviewer; the mothers' ratings of smileyness are not related to household socioeconomic status or other maternal characteristics. Children mothers rate as shy make poorer impressions on the interviewer, have mothers who have rated themselves as shy, and show somewhat more behavioral problems; again, relationships with socioeconomic status are sparse.

For the remaining scales--fearful-fussy, compliant, and dependent-demanding, there is evidence for construct validity as well as evidence that

ratings are associated with household and maternal characteristics. Children described as less fearful-fussy have mothers who are less shy and they make better impressions on the interviewers; they also show better cognitive performance, even with interviewer impressions statistically controlled. Mothers from less advantaged households rate their children as more fearful-fussy than mothers from more advantaged households. Similarly, children described as compliant make a good impression, perform well on the vocabulary assessment, and show fewer behavioral problems; and they come from households richer in material and human resources. The pattern for dependent-demanding is the same: children perceived as more dependent-demanding are more apt to live in socioeconomically disadvantaged households; they make a worse impression on the interviewer, perform worse in the vocabulary assessment, and have more behavior problems.

The variability in the relationships of our temperament scales and household context variables provides some evidence regarding discriminant validity. The weak pattern of relationships between context and dimensions of infant temperament supports the notion of temperament heritability for measures such as smileyness and shyness, it also reflects the fact that environmental influences have not been at work long enough to have effects discernable in survey measurement. The tendency of lower socioeconomic status mothers to perceive their infants as more active and less predictable foreshadows the more consistent relationships between temperament and socioeconomic status among older children. For older children, maternal-child interaction has been ongoing long enough for context to have affected both the expression of child temperament and maternal perceptions of temperamentally relevant child behaviors. Thus, consistent relationships between indicators of maternal socioeconomic status and maternal perceptions of child temperament likely reflect the child's actual temperament, maternal perceptions of it, and household contextual influences on behavioral expression of temperamental tendencies.

We have argued that the Child survey attached to the mothers of the NLSY provides a potentially important resource for researchers interested in child development. We believe we have enhanced this potential by deriving several reliable measures of maternal perceptions of child temperament that also have demonstrated construct validity. We have examined their relationship to maternal contextual variables, and identified those scales most affected by maternal context. The potential now exists for additional substantive research to further investigate the behavior of these measures in multivariate analysis.

It is important to acknowledge, however, limitations of this resource for investigators to consider as they construct their research agendas. First, many developmental psychologists appear most comfortable working with data they have collected themselves. Under these conditions, they are in an excellent position to verify the accuracy of their measurements and to generally "be in touch" with their data. This is not possible in any secondary analysis context. While many sociologists appear willing to accept this limitation if other advantages are present, not all researchers will agree with this choice. They may argue that in the case of child development, it is critical for researchers to have the degree of direct knowledge of their subjects that only primary data collection can afford. However, others may be willing to trade such direct knowledge in exchange for the advantages that this data set allows. We have argued that large sample sizes open the door to techniques of parameter estimation that may not be used when sample sizes are below 200. We have also argued that patterns of statistical interaction can be displayed in large samples in a way that is not possible in many smaller samples.

A second limitation may also act as a collective advantage. This limitation is again inherent in the secondary nature of the data: that the investigator will have no control over the questions asked of the respondents. In a primary data collection situation, investigators have this control, but this leads to differences in measuring strategies across investigations, thus

hampering cumulation of findings across investigators. Assuming a number of investigators work with this data set over time, there will be fewer variations in measurement strategies, which should contribute to the cumulation of findings across investigators. We believe the analyses we have presented here will facilitate such cumulation by delineating the measurement properties of scales that can be derived from the data. We look forward to substantive research involving these measures as a mechanism to further assess scale validities as well as for the insights such research will provide on child development.

References

- Bates, J. E. 1980. "The Concept of Difficult Temperament." *Merrill-Palmer Quarterly* 26(4):299-319.
- Bates, J. E., C. Freeland, and M. Lounsbury. 1979. "Measures of Infant Difficultness." *Child Development* 50:794-803.
- Cameron, J. 1977. "Parental Treatment, Children's Temperament and the Risk of Childhood Behavior Problems, I." *American Journal of Orthopsychiatry* 47:140-147.
- Campos, J. J., K. Barrett, M. E. Lamb, H. H. Goldsmith, and C. Stenberg. 1983. "Socioemotional Development." Pp. 783-915 in *Handbook Of Child Psychology, Volume 2: Infancy and Developmental Psychobiology*, edited by M. M. Haith and J. J. Campos. New York: Wiley.
- Carey, W. B., and S. McDevitt. 1977. "Temperament As A Factor In Early School Adjustment." *Pediatrics* 60:621-624.
- Costa, P. T., Jr., R. R. McCrae, and D. Arenberg. 1980. "Enduring Dispositions in Adult Males." *Journal of Personality and Social Psychology* 38:793-800.
- Goldsmith, H. H., A. Buss, R. Plomin, M. K. Rothbart, A. Thomas, S. Chess, R. Hinde, and R. McCall. 1987. "Roundtable: What Is Temperament? Four Approaches." *Child Development* 58:505-529.
- Kagan, J., J. S. Reznick, C. Clarke, N. Snidman, and C. Garcia-Coll. 1984. "Behavioral Inhibition to the Unfamiliar." *Child Development* 55:2212-2225.
- Korn, S. J. 1984. "Continuities and Discontinuities in Difficult/Easy Temperament: Infancy to Young Adulthood." *Merrill-Palmer Quarterly* 30(2):189-199.
- Matheny, A. P., Jr., R. S. Wilson, and A. S. Thoben. 1987. "Home and Mother: Relations With Infant Temperament." *Developmental Psychology* 23(3):323-331.
- National Longitudinal Survey Handbook*. 1987. Columbus, Ohio: Center for Human Resource Research, The Ohio State University.
- Parcel, T. L., and E. G. Menaghan. 1988. "Measuring Behavioral Problems in A Large Cross Sectional Survey: Reliability and Validity For Children of the NLS Youth." Unpublished manuscript.
- Rothbart, M. K., and D. Derryberry. 1984. "Emotion, Attention and Temperament." in *Emotion, Cognition and Behavior*, edited by C. E. Izard, J. Kagan and R. Zajonc. New York: Cambridge University Press.
- Ruopp, R. and J. Travers. 1982. "Janus Faces Day Care: Perspectives on Quality and Cost." Pp. 72-101 in *Day Care: Scientific and Social Policy Issues*, edited by E. Zigler and E. W. Gordon. Boston: Auburn House.
- Vaughn, B. E., C. F. Bradley, L. S. Joffe, R. Seifer, and P. Barglow. 1987.

"Maternal Characteristics Measured Prenatally Are Predictive of Ratings of Temperamental 'Difficulty' on the Carey Infant Temperament Questionnaire." *Developmental Psychology* 23(1):152-161.

Waters, E., and K. E. Deane. 1985. "Defining and Assessing Individual Differences In Attachment Relationships: Q-Methodology and the Organization of Behavior in Infancy and Early Childhood." *Monographs of the Society For Research in Child Development* 50(1):41-65.

Table 1. Cognitive, Socio-Emotional, and Developmental Environment Measures for Children of NLSY Women.

<u>Cognitive Measures:</u>	<u>Ages of Children Assessed</u>
1. Weschler Intelligence Scale for Children Complete Digit Span Subscale	7 years and older
2. Peabody Individual Achievement Test Complete Math Subscale Complete Reading Recognition Subscale Complete Reading Comprehension Subscale	5 years and older
3. Peabody Picture Vocabulary Test Complete Test	3 years and older
4. McCarthy Scales of Children's Abilities Complete Verbal Memory Subscale	3 years through 6 years, 11 months
5. Memory for Locations (Jerome Kagan) ^a	8 months through 3 years, 11 months
6. Body Parts (Jerome Kagan)	1 year through 2 years, 11 months
<u>Socio-Emotional Measures</u>	
7. Perceived Competence Scale for Children (Susan Harter) Complete General Self-Worth Subscale Complete Academic Ability SubScale	8 years and older
8. Achenbach Child Behavior Checklist Subset of 28 items; mothers' reports	4 years and older
9. Temperament/Behavioral Style Maternal report items vary by child's age; 6 items apply only to infants, 11 to children under 2, and 20 items to children 2 years and older. Seven interviewer rating items apply to children 8 months and older.	Newborn through 6 years, 11 months
10. Motor and Social Development Scale (Gail Poe)	Newborn through 3 years, 11 months
<u>Developmental Environment</u>	
11. Home Observation for Measurement of the Environment (Caldwell and Bradley) Subsets of items from the Infant, Preschool, and Elementary versions of the HOME scale; include mothers' reports and interviewer ratings. Items vary by children's age.	All Children

^aOriginators of measures are indicated in parentheses.

Table 2. Factor-Based Infant-Toddler Temperament Scales: Factor Structures and Item Statistics.

ITEM CONTENT	FACTOR-ITEM CORRELATION ^d				MEAN ^e	S.D.
	Infant Items		Infant-Toddler Items			
	I	II	I	II		
<u>Babies Under AGE 1:</u>						
Squirms and kicks during feeding ^a	.679				2.65	1.36
Waves arms during feeding ^a	.794				3.28	1.33
Moves in crib during sleep ^a	.437				2.74	1.26
Baby is sleepy at same time each P.M. ^a		.722			4.08	1.09
Baby is hungry at same times each day ^a		.699			4.08	1.05
Consistency of mood at A.M. wake up ^a		.437			4.54	.80
<u>Babies and Toddlers under Age 2:</u>						
Cry when seeing stranger ^a			.583		2.16	1.36
Cry when seeing strange dog/cat ^a			.437		1.62	1.10
Cry when you leave room ^a			.428		2.51	1.28
Cry at the doctor ^a			.569		2.43	1.51
Cry at loud, unexpected sound ^a			.392		2.09	1.24
Difficulty calming child ^b if upset ^a			.402		1.70	.98
Frequency of fussy times ^b			.369		2.48	.82
General quantity of fussiness ^c			.441		2.33	.78
Smile when you play ^a				-.713	4.73	.75
Smile when s/he plays alone ^a				-.695	3.87	1.06
Smile when in bath ^a				-.662	4.23	1.17
Eigenvalues	1.88	1.70	2.45	2.20		
Percent of Variance	31.3	28.3	22.3	20.0		
Factor-Based Scales:						
Number of Items	3	3	8	3		
Cronbach's alpha	.66	.64	.67	.75		

^a Responses to these questions are in terms of the frequency this activity is observed. Items were scored 1 (almost never), 2 (less than 1/2 the time), 3 (1/2 the time), 4 (more than 1/2 the time), 5 (almost always).

^b Responses to this question were coded as 1 (almost never), 2 (once or twice a day) 3, (couple times in AM and PM), 4 (several times a day), 5 (almost every hour).

^c Responses to this question were coded as 1 (almost never), 2 (less than average), 3 (about average), 4 (more than average), 5 (almost always).

^d All factor-item correlations .35 and higher are displayed. Factor-analytic solutions were estimated using weighted data, and specified principal axis factoring and oblique rotation. Due to listwise deletion, actual number of cases used in the factor analysis for infants through 11 months of age was 560; actual number of cases used in the second factor analysis, for infants and toddlers through 23 months of age, was 1,139. Factor labels for infants under 1 year are I, Active; II, Predictable; labels for factors derived for babies and toddlers under 2 are I Fearful-Fussy; II, Smiley.

^e Item statistics were calculated with weighted data. Actual number of cases is 560 for the first six items, and 1,139 for the remaining 11 items.

Table 3. Factor-Based Toddler-Child Temperament Scales: Factor Structure and Item Statistics.

ITEM CONTENT	FACTOR-ITEM CORRELATION ^b			MEAN ^c	S.D.
	I	II	III		
Eat what is served ^a	-.397			4.03	1.02
After protest, obey and eat ^a	-.465			3.65	1.14
Resist bedtime ^a				2.62	1.43
After protest, obey and go to bed ^a	-.501			3.83	1.33
Turn off TV without protest ^a	-.468			3.87	1.39
After protest, turn off TV ^a	-.583			4.17	1.32
Share toys with other children ^a	-.423			3.87	1.03
Cry when hurts self a bit ^a			.403	2.78	1.27
Smile/laugh when alone ^a				4.22	.96
Aggress against other children ^a			.433	2.01	1.04
Hard for mom to soothe child ^a			.374	1.84	1.09
When playing, child stays close ^a			.416	2.79	1.41
Child imitates mother ^a				3.18	1.27
Upset if mom leaves and child alone ^a			.548	1.64	1.06
Child demanding when mom is busy ^a			.562	2.67	1.19
If mom upset, child tries to help ^a				3.84	1.21
Child wants mom's help ^a			.346	3.36	1.19
Shy in meeting new child ^a		.507		2.35	1.42
Shy in meeting new adult ^a		.564		3.13	1.45
Child acts shy initially (Int) ^d		.660		2.08	1.11
Child acts shy at end of interview (Int) ^d		.632		1.84	1.00
<hr/>					
Eigenvalues	3.29	2.06	1.79		
Percent of Variance	15.7	9.8	8.5		
<hr/>					
Factor-Based Scales:					
Number of Items	6	4	7		
Cronbach's alpha	.64	.68	.63		

^a Responses to these questions are in terms of the frequency this activity is observed. Items were scored 1 (almost never), 2 (less than 1/2 the time), 3 (1/2 the time), 4 (more than 1/2 the time), 5 (almost always).

^b All factor-item correlations .35 and higher are displayed. Factor-analytic solutions were estimated using weighted data, and specified principal axis factoring and oblique rotation. Actual number of cases used in the factor analysis was 2,433, due to listwise deletion of cases with any missing values. Factor labels are I, Compliant; II, Shy; III, Dependent-Demanding.

^c Item statistics for these twenty one items are calculated using weighted data and are reported for the 2,433 children between 2 years and six years eleven months who had valid responses on all the items.

^d These items were coded using a one to five scale where one represented not at all shy and anxious, or sociable and friendly and five represented extremely shy/quiet/withdrawn.

Table 4. Factor-Based Scales Based on Interviewer Assessments:
Factor Structure and Item Statistics.

ITEM ^a	FACTOR-ITEM CORRELATION ^b		MEAN ^c		S.D.	
	8-23 Mos. I	24-83 Mos. I	8-23 Mos.	24-83 Mos.	8-23 Mos.	24-83 Mos.
Attitude toward being tested	.89	.90	2.98	3.77	1.07	1.09
Rapport with interviewer	.72	.85	3.32	3.88	1.01	.99
Perseverance/ persistence	.91	.90	2.88	3.64	1.19	1.12
Cooperation	.92	.91	2.87	3.77	1.19	1.10
Motivation/ interest	.91	.92	2.90	3.68	1.17	1.11
Eigenvalues	4.02	4.22				
Explained Variance	80.4	84.4				
Factor-Based Scales:						
Number of Items	5	5				
Cronbach's alpha	.94	.95				

^a These items were coded on a one to five scale where one represented poor, three represented average and five represented excellent.

^b All factor-item correlations .35 and higher are displayed. Factor-analytic solutions were estimated using weighted data, and specified principal axis factoring and oblique rotation. Due to the listwise deletion of cases with any missing values, actual number of cases used in the factor analyses was 790, for babies and toddlers 8 to 23 months, and 2,609 for children 24 to 83 months. Factor labels for both solutions are I, Positive Impression.

^c Item statistics for the interviewer assessment items are reported for the 790 children between 8 and 23 months and the 2,609 children between 24 and 83 months, who had valid responses on the interviewer items.

Table 5. Evidence for Construct Validity of Temperament Scales: Partial Correlations with Cognitive Assessments, Interviewer Impressions, and Maternal Temperament, Controlling for Child Age.

<u>Babies under 1 Year:</u>	ACTIVE		PREDICTABLE
Interviewer Impression ^a	-.13ns		.15
Memory for Location ^b	-.12ns		.01ns
Maternal Shyness	.05ns		-.06ns
<u>Babies and Toddlers under Age Two:</u>	FEARFUL-FUSSY		SMILEY
Interviewer Impression	-.16		.08
Memory for Location ^b	-.15		-.02ns
Maternal Shyness	.10		-.07
<u>Children Aged Two through Six:</u>	COMPLIANT	SHY	DEPENDENT-DEMANDING
Interviewer Impression	.11	-.48	-.17
Memory for Location ^b	.03ns	.01ns	.02ns
PPVT ^c	.18	-.14	-.25
Maternal Shyness	-.04	.13	.03ns
Behavior Problems ^d	-.28	.09	.46

^a Interviewer impressions are asked only for children 8 months and older.

^b Memory for location is asked only for children between 8 months and 3 years, 11 months.

^c PPVT is asked only for children at least three years of age.

^d Behavior Problems is a factor based composite measured for children 4 years and older (see Parcel and Menaghan, 1988).

Table 6. Partial Correlations between Temperament and Maternal Background Characteristics, Controlling for Child Age: Factor-Based Scales for Babies and Toddlers Under Age Two.

FACTOR-BASED SCALES:	ACTIVE	PREDICT	FEARFUL- FUSSY	SMILEY
<u>Current Economic and Occupational Conditions</u>				
Household Income	-.12	.10	-.11	-.00ns
Poverty Status (1=yes)	.04ns	-.15	.15	-.03ns
Mother's Employment Status (1=employed)	-.02ns	-.00ns	-.07	.00ns
Mother's Marital Status (1=married)	.02ns	.09	-.13	.01ns
If Mother is Employed:				
Mother's Hourly Rate of Pay	-.04ns	.02ns	.02ns	-.04ns
Mother's Occupational Status	-.05ns	.01ns	-.11	-.07ns
<u>Family Background Characteristics</u>				
Mother's Educational Attainment	-.12	.06ns	-.13	.01ns
Mother's Cognitive Achievement	-.19	.20	-.26	.02ns
Minority Status (1=non-white)	.13	-.25	.22	-.05ns
<u>Child Characteristics</u>				
Gender (1=male)	.00ns	.05ns	.02ns	-.07
Birth Order	.12	-.04ns	.24	-.08

The mother's cognitive achievement is measured by the Armed Services Vocational Aptitude Battery, administered in 1980. All other measures are as reported in 1986.

Table 7. Partial Correlations Between Temperament Variables and Maternal Background Characteristics, Controlling for Child Age, Factor-Based Scales for Children Age Two through Six.

FACTOR-BASED SCALES:	COMPLIANT	SHY	DEPENDENT- DEMANDING
<u>Current Economic and Occupational Conditions</u>			
Household Income	.09	.04ns	-.17
Poverty Status	-.07	.01ns	.16
Mother's Employment Status (1=employed)	.02ns	.01ns	-.05
Mother's Marital Status(1=married)	.08	.04	-.15
If Mother is Employed:			
Mother's Hourly Rate of Pay	.05	-.02ns	-.08
Mother's Occupational Status	.06	-.07	-.06
<u>Family Background Characteristics</u>			
Mother's Educational Attainment	.06	-.04	-.09
Mother's Cognitive Achievement	.20	-.04ns	-.34
Minority Status (1=nonwhite)	-.13	-.01ns	.22
<u>Child Characteristics</u>			
Gender (1=male)	-.05	.01ns	-.01ns
Birth Order	.02ns	.13	.06

Note. Correlations are calculated with weighted data, using pairwise deletion for all cases with valid scores on all temperament measures; minimum unweighted number of cases is 2,433. All correlations are statistically significant at the .05 level or better, unless flagged with ns (not significant). Correlations using mother's hourly rate of pay and the Duncan measure of the status of her occupation are calculated for employed mothers only.

The mother's cognitive achievement is measured by the Armed Services Vocational Aptitude Battery, administered in 1980. All other measures are as reported in 1986.