

9/88

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

By

Toby L. Parcel and Elizabeth G. Menaghan

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 BEHAVIORAL PROBLEMS 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 samples of fewer than 200 children. Studies reporting findings from samples of 30 to 60 are not uncommon. Although replication of findings contributes some evidence for external validity, these investigations cannot entirely overcome limitations inherent in small samples. Small samples often preclude investigators from estimating parameters portraying determinants of child outcomes, and thus investigators rely more heavily on statistical tests pointing to the differences between or among groups of children, possibly controlling for a few covariates. In addition, small samples cannot be appropriately divided to portray the dynamics of statistical interaction, even when literature suggests that understanding the magnitude of such interaction may be critical to theory testing. 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 5,876 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. Since

we cannot assume that subsets of scales have identical measurement properties to those of the original measures, this study reports the first investigation of the measurement properties of one of these measures, a subset of items from the Child Behavior Checklist developed by Thomas Achenbach (1978). 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 outcomes 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 scales must have acceptable levels of reliability and demonstrated validity. In this paper, we develop measures of children's behavioral problems and provide preliminary evidence regarding their validity.

SCALE CONSTRUCTION AND RELIABILITY OF MEASURES

Measuring Children's Behavioral Problems

Classifying children with reference to existence and severity of behavior problems is of major concern to clinicians and researchers. Certainly measurement of behavioral problems is an important child outcome in its own right, as well as one that may be predictive of subsequent social adjustment (Mechanic, 1980; Caspi and Elder, 1987). Perhaps the most widely used and best documented measure of children's behavioral problems and competencies is the Child Behavior Checklist developed by Thomas Achenbach and Craig Edelbrock (1983). In development and use since the mid-1960's, the CBCL has gone through several major revisions, been normed and validated on relatively large clinical and normative samples, and been utilized in many studies (see Achenbach and Edelbrock, 1978, for a review). Parents of children aged four through sixteen are asked to respond to 118 items about behavior problems, and to answer twenty questions about specific child activities and competencies.

Each of the original items in the CBCL had been shown to discriminate between children referred for treatment for behavioral problems and those not referred, even when gender, age and socioeconomic status were statistically controlled (Achenbach and Edelbrock, 1983). Under guidance from the NICHD panel, 28 items from the 118 behavior problem items were selected for inclusion in the survey. Items that tapped rare behavioral problems, unlikely

to be present with appreciable frequency in a non-clinical sample, were excluded from consideration. Following Peterson and Zill (1986), items were selected that had high loadings on clinically significant factors derived from factor analyses of the entire 118 item scale, such as aggressiveness, depression, and hyperactivity. Items were selected so as to try to capture the diversity of behavioral problems evident in a sample of children.

Our expectations for the exploratory factor analyses were guided by three possibilities. First, we expected that data reduction would suggest that the 28 items could be adequately explained by a smaller number of underlying factors. Work by Achenbach and Edelbrock (1983) and Peterson and Zill (1986) provides support for this expectation. Achenbach and Edelbrock (1983) performed a series of factor analyses on the entire 118 item Checklist by gender for children aged 4 to 5, 6 to 11 and 12 to 16. Although the precise solutions varied for the different subgroups, factors labeled "Aggressive", "Hyperactive," "Immature," "Depressed," "Delinquent," and "Somatic Complaints" were derived across several solutions. While the initial solutions were derived from analyses of clinical samples, the findings were normed against a general sample of children aged 4 to 16; validity was established via comparisons between scores from the clinical and non-clinical samples. Peterson and Zill (1986) constructed scales of behavior problems from a subset of CBCL items that were included in the 1981 National Survey of Children. This subset is quite similar to that used for the NLSY survey. Peterson and Zill (1986) suggest these scales represent depressed/withdrawn behavior, antisocial behavior and impulsive/hyperactive behavior. Their research suggests that marital disruption is associated with several behavior problems for children, although living arrangements after disruption and marital history condition these relationships. Thus there is sufficient evidence to suggest the validity of submeasures derived from data reduction performed on

the entire set of 28 items.

Second, we anticipated the possibility that these items could be grouped into a smaller number of broad band groupings of behavior problems than would be directly suggested by the factor solutions noted above. Achenbach and Edelbrock (1983) note that clinical child researchers have repeatedly identified two broad-band groupings of behavior problems, reflecting a distinction between fearful, inhibited, over-controlled behavior and aggressive, antisocial, and undercontrolled behavior; they suggest that these two more general constructs, which they label Internalizing and Externalizing, may account for the myriad of specific behaviors tapped by their Checklist. Their empirical evidence is consistent with these ideas.

Third, we anticipated the possibility that the items selected might be sufficiently highly correlated so that a single factor might adequately account for data variation. It is more difficult to discriminate clinically significant syndromes among normative samples than in clinical samples where behavior problems appear more sharply discriminated from one another (Achenbach and Edelbrock, 1983). In normative samples, behavior problems tend to be associated; while manifestations of behavioral problems *may* vary across children, researchers have frequently observed that externalizing and internalizing behaviors appear in the same child. Therefore, the items may detect the presence or absence of problems generally, independent of their particular behavioral manifestation. Just as there tends to be a correlation between verbal and analytic abilities in the measurement of cognitive outcomes, we expect a positive relationship between the two broad band groups of behavioral problems and positive correlations among most of the original items, particularly since our sample is a non-clinical one.

METHOD AND RESULTS

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 precision of the estimates derived using the data, particularly if one is interested in estimates for such subgroups.

Mothers were asked to rate the behavior problems of each child four years of age or older. They responded in terms of Achenbach's zero to two scale where zero represented that the item was not true of their child, 1 represented that the item was somewhat or sometimes true and 2 that the item was very true or often true. Our initial analyses used as respondents the 1,792 children with valid responses on the complete set of 28 items. These 28 items, their means and standard deviations, and the loadings from the factor solution described below are listed in Table 2. Variation in the levels of means across the items suggests that relatively few mothers agree that their children appear to feel worthless, act withdrawn, destroy things or are unliked by other children, while many more agree that their children argue too much, are restless and have sudden changes in mood.

We performed a series of factor analyses to evaluate the number and configuration of factors underlying the data. We first subjected the included items to a factor analysis using principal axis factoring and an oblique rotation. Five factors met the eigenvalue criterion and all were

substantively interpretable. We interpreted an item as loading on a factor if it loaded at an absolute level of .40 or greater. The first factor we have named Peer Problems because loading on it are items representing disobeys at school, not liked by peers, fails to get along with peers, destroys, bullies, cruel or mean, cheats, disobeys at home, is in conflict with teachers and is unhappy. The second factor is named Aggressive; loading on it are fails to get along with peers, bullies, cheats, disobeys at home, has a bad temper, is stubborn, argues and is moody. The third factor we have labeled Hyperactive. Loading on it are items representing fearful, confused, tense, poor attention, impulsive, obsessed, restless, cheats and is stubborn. The fourth factor we have named Depressed. Items loading on it are feels unloved, worthless, unhappy, withdrawn, fearful and confused. Items loading on the fifth factor are clings to adults, cries too much, demands attention, and is too dependent. We name it Whine; it is most similar to what Achenbach and Edelbrock (1983) label Immature. We formed factor-based scales corresponding to the factor solution described above. Items were transformed into Z scores prior to creating factor-based scores. Alpha reliabilities for the composites are: Peer Problems: .80; Aggressive: .80; Hyperactive: .80; Depressed: .73; Whine: .64.

In evaluating our second and third expectations, we found several pieces of information suggesting that a smaller number of factors might adequately account for the data. The pattern of overlapping factor loadings as shown in Table 2 suggests overlap between the factor we call Aggressive and the factor we call Peer Problems, as well as between Hyperactive and Depressed, lending credence to the notion that two dimensions or even a single factor might describe the data adequately.

Table 3 provides two types of evidence to help us evaluate this idea more precisely. The lower triangle of the table contains the correlation matrix

among the factors directly derived from the factor analytic solution. In general, the correlations among the five factors are moderate to moderately strong. There are correlations in the .38 to .44 range among the factors we named Aggressive, Hyperactive and Peer Problems, and somewhat lower correlations between each of these factors and Depressed. In support of the idea that behavior problems generally will be interrelated, Whine correlates positively with Aggressive, Hyperactive and Peer Problems. This finding is consistent with findings from Achenbach and Edelbrock (1983) who found that dimensions such as these tend to be more positively intercorrelated in their normative samples than in their clinical samples. Children who have been clinically diagnosed with behavioral problems may have more easily discriminable clusters of behaviors, thus leading to lower intercorrelations.

The upper triangle of the matrix contains the correlations among our factor based composites. The correlations are even stronger than those among the factors, and provide additional evidence that a smaller number of underlying constructs may explain the observed correlations. While we can attribute in part the relationships among the scales to our use of overlapping items across the measures, this decision does not completely account for the findings. In additional analysis (not presented here) we reduced factor complexity by examining the items loading on more than one factor, and removing them from the factor on which the loading was clearly lower; this left only three items loading on more than one factor. We re-calculated correlations among these revised composites. Correlations remained high under these conditions. While the correlations among the three factors we have classified as representing externalizing behavior patterns are the strongest, the remaining relationships are also sizeable. It appears, then, that while particular factor based scores might conceptually correspond to the externalizing-internalizing distinction, pervasiveness of moderate to strong

correlations among the factors and scales also supports the idea of a single underlying factor explaining the data.

We performed additional analysis to evaluate whether broad band groupings of behaviors could adequately describe the data as well as whether a single underlying factor could adequately account for factor variation. We performed a higher order factor analysis where the input matrix consisted of the correlations among the factors derived in the five factors solution described above. We forced the solution to extract two factors and performed both oblique and orthogonal rotations. Table 4 presents the orthogonal solution that suggests the reasonableness of viewing the items in terms of Externalizing and Internalizing. Aggressive, Peer Problems and Hyperactive clearly load on the factor we have named Externalizing, while Depressed and Whine load on the factor we name Internalizing.

Three pieces of evidence, however, suggest that a single factor might adequately account for factor variation. First, in the forced two factor solution described above, only the first factor had an eigenvalue of over 1.00, suggesting that it is not appropriate to interpret the second factor. Second, the oblique rotation of this solution (not presented here) shows similar loadings across the two higher order factors for several input factors and a high correlation ($r = -.68$) between the two extracted factors. Third, even in the orthogonal solution, Whine loads above the .40 criterion on both factors. Given this evidence, we performed a second higher order factor solution that does not constrain the number of factors to two. These results are summarized in the third panel of Table 4. We see that a single factor solution adequately accounts for variation among the first order factors, and that loadings on the input factors all exceed an absolute .40.

On the basis of these findings we constructed three additional measures. Because of the potential usefulness of measures for the broad band groupings

of externalizing and internalizing, we constructed an Externalizing and an Internalizing scale where each was composed of the nonredundant items associated with Aggressive, Peer Problems, and Hyperactive for the Externalizing scale, and Depressed and Whine for Internalizing. Three items (unhappy, confused, and fearful) load on both broad band factors. The reliability of the Externalizing scale is .88 and the reliability of the internalizing scale is .77. The correlation between these two scales is .73, a finding consistent with the second higher order factor solution just discussed. For the third measure, we also constructed an overall measure of behavior problems by summing the Z score for each of the 28 items. The reliability of this factor based scale is .90, suggesting the measure has high reliability as a general measure of behavior problems.

These findings suggest that three sets of measures at differing levels of generality can be derived from the original 28 items contained on the Child survey. If an investigator wants to discriminate among types of behavioral problems at some level of detail, the first factor solution described above suggests five measures with adequate to good reliabilities that might prove useful in a number of research contexts. If the purpose of the research is such that only a single discrimination between undercontrolled and overcontrolled behaviors is useful, the factor based scales of Externalizing and Internalizing are appropriate choices. Finally, for investigators who do not wish to discriminate between or among the more detailed behaviors, but rather need a general measure of behavior problems, possibly for control purposes, the measure based on the final factor solution would be appropriate.

VALIDITY

In order for the research community to use these measures with confidence, we must be assured that there is evidence for their validity. While evidence regarding measurement validity will accumulate as the measures

are used in research, we provide initial evidence consistent with the notion that the factor based scales we have constructed have construct validity.

Achenbach and Edelbrock (1983) report that lower socioeconomic status parents report fewer child competencies and more problems than upper SES parents, but that the effects of SES on problems scales are small. They also found few effects of race. Still, given the pervasive influence of socioeconomic status and race on socialization of children (Gecas, 1979; Peters and Massey, 1983), we expect that there will be discernible relationships between our scales and measures of these constructs, or between our scales and other socio-demographic variables associated with race or class, such as measured mental ability and maternal schooling. We expected that measures of behavioral problems would correlate with several measures of current economic and occupational conditions including maternal annual earnings, household income, maternal annual rate of pay, and maternal occupational status as measured by the Duncan SEI. We expected a positive relationship with poverty status (coded 1 = in poverty, 0 otherwise).

We also expected negative correlations between children's behavioral problems and maternal background characteristics including years of schooling, maternal cognitive performance (assessed in 1980 for all NLSY respondents with the Armed Services Vocational Aptitude Battery), maternal age at birth of this child, and maternal marital status (Hernandez, 1986). We expected that maternal adult shyness would correlate positively with behavior problems, particularly with measures of internalizing behaviors. Finally, we expected relationships with child characteristics. Following work by Achenbach and Edelbrock (1983), we expect boys to be rated more highly on the Externalizing scale and its components and girls to be rated more highly on the Internalizing Scale and its components. Since minority status is related to more impoverished neighborhood and household environments, we also expect

Black and Hispanic children to be rated more highly on measures of behavioral problems than non-Black and non-Hispanic children. We also expect age differences such that older children should be more likely to exhibit Externalizing behavior. Predictions for age differences in Internalizing behaviors vary: while younger children should be more likely to exhibit some Internalizing behavior, particularly whining, crying, and clinging, we expect older children to show greater levels of depressed behavior than younger children. We expect that first born children will have fewer behavioral problems. Finally, we expect that behavior problems and cognitive competencies will be associated for two reasons: behavior problems may interfere with the acquisition of cognitive skills, and frustrating experiences in school settings may impair self-confidence and self-control. Thus, we expect an inverse association between behavior problems and cognitive skills as assessed by the Peabody Picture Vocabulary Test.

Table 5 displays a series of bivariate correlations between the measures of the constructs that we had theoretical reason to believe would be associated with the respective scales. The findings suggest that these expectations are generally supported. There are consistent negative relationships between household income and all measures of behavior problems, and consistent positive relationships of all measures with poverty status. Mothers who are married, and thus in households with greater economic and human resources, are less likely to report behavior problems than unmarried mothers. The total amount of mothers' earnings is inversely correlated with global measures of both Internalizing and Externalizing problems, but is significantly associated only with Peer Problems and Aggressive behavior among the subscales.

Employed mothers are less likely to report behavioral problems than not-employed mothers. Among employed mothers, occupational status is not significantly associated with behavioral problems; but the higher the mother's hourly rate of pay, the lower the level of externalizing behavior problems. There is no relationship between rate of pay and the internalizing measures. This is one of the few correlates that shows a systematic pattern of discrimination between Externalizing and Internalizing.

Turning to family background characteristics, the measures are generally negatively correlated with mother's educational attainment as well as with education of mother's spouse. Mothers who bore the child at an older age, and those who had higher cognitive aptitude scores, also report fewer behavior problems. The only significant association with maternal adult shyness is Depressed. Finally, minority status (which contrasts Black and Hispanic children with all others) is generally not significantly associated with reported behavior problems.

Concerning child characteristics, older children are somewhat more likely to score higher on the problems measures, except Whine where the relationship is negative and significant. The opposite age trends for Depressed and Whine produce a non-significant age association for the Internalizing measure. Males are scored more highly than females on all externalizing measures. The two subscales of Internalizing differ somewhat; there are no gender differences on Depressed, but girls score higher on Whine and thus somewhat higher on the overall Internalizing measure. Later born children are more likely to score highly on all measures except Depressed. Finally, there are relatively strong negative relationships between children's cognitive vocabulary knowledge (PPVT) and all measures of behavior problems.

As expected, those children who have obtained professional help for psychological problems score noticeably higher than children who have not had

such help on all the measures. That the relationship is weakest for Whine suggests that this behavior by itself is less likely to motivate obtaining professional help than are the behaviors associated with the Externalizing composite. There are also significant positive associations between being medicated for behavioral problems and Aggress, Hyperactive, Externalizing, Internalizing and Total, although not for Peer Problems, Depressed or Whine.

While the bivariate associations are low to moderate, they are consistent with theoretical expectations and previous findings. Externalizing and its subscales behave similarly; and the findings for the Total Problems measure, which is disproportionately composed of Externalizing items, are also consistent with the overall picture for the Externalizing measures. The two subscales of Internalizing differ somewhat, with older children and those with shy mothers more likely to be viewed as being depressed, but younger children and girls more apt to be characterized as whiners.

We can provide additional evidence regarding the construct validity of these measures by studying patterns of statistical interaction for each involving race, gender, age and mother's marital status. Bronfenbrenner, Alvarez and Henderson (1984) and Hoffman and Nye (1974) provide relevant arguments regarding gender-specific age periods of greatest developmental risk. Bronfenbrenner et al. (1984) suggest that the period of greatest risk for boys is in the preschool years, while girls are most vulnerable during adolescence. One major stress during childhood is disruption due to divorce, and we consider the possibility that these processes may operate differentially by race. Thus, for each measure we study variation by race, gender, age and mother's marital status. We distinguish Hispanics and Blacks from others in assessing race; we categorize age as 4 and 5 year olds; 6, 7, and 8 year olds; and 9 years old and older. Mother's marital status is categorized as never married, currently married with spouse present, and formerly married. Table 6

reports four way analyses of variance for Externalizing, Internalizing and Total. The findings are generally similar for both Externalizing and Internalizing problems. Race is not significant as a main effect nor as an interactive effect. Gender and mother's marital status are significant as main effects and contribute to interaction, while age is not significant as a main effect but contributes to patterns of interaction.

Figure 1 provides a visual summary of the findings. Our findings show that preschool boys of formerly married mothers score the highest on the Externalizing scale, with same age boys of married mothers scoring the lowest and boys of never married mothers intermediate. These differences decline as the boys get older and are minimal for the nine years and older age group. Among girls, we find practically no differences in Externalizing in the preschool years regardless of maternal marital status, but these differences increase over time reaching their greatest in the nine years and over age group.

The findings regarding Internalizing are of similar form but lesser magnitude. Preschool boys show noticeable differences in Internalizing as a function of maternal marital status, with boys of married mothers scoring lower on Internalizing than boys of formerly married mothers; these differences disappear in the oldest age group. Among girls, differences in Internalizing are minimal during the preschool years but are more noticeable for school age and older girls. The patterns for Total Problems, not displayed in Figure 1, are more similar to those for Externalizing than Internalizing since Externalizing contributes more heavily to its construction. Additional analyses (not presented here) were also conducted for each of the five scales derived in the first factor analysis. Findings for the narrower measures were generally consistent with these presented here with minor variations introduced as a function of the more specific nature of

those composites. These findings are consistent with our expectations regarding the interactive effects of age, gender, and maternal marital status on behavior problems, although our expectations regarding race were not supported.

MODIFYING MEASURES FOR YOUNG CHILDREN

Although Achenbach and Edelbrock developed the CBCL for use with children aged four through sixteen, the inclusion of items asking about school problems introduces additional complications to scale construction. In the abridged set of 28 items used for the NLSY children, two (asking about relationships with teachers and disobedience in school) were asked only for children in school or preschool since they pertained specifically to the school setting. Since many of the four and five year old children were not attending preschool or kindergarten, they had missing values on these items and were omitted from initial factor analyses, scale construction, and validity assessment. This resulted in the unweighted N of 1,792 used in the above analysis. Researchers interested in using this data set, however, would be loath to omit the many four and five years olds who had valid answers on all but these two items. Construction of measures for children not yet attending school would have to omit these two items.

To assess the possible implications of omitting these two items, we constructed modified measures for the scales that included these two items (the Peer Problems and Hyperactive subscales, the Externalizing measure, and the Total Problems scale). Next, we assessed the reliabilities for these restricted measures for the larger pool (unweighted N=2,673) of children with valid scores on the 26 non-school-related items. These reliabilities, displayed in Table 7, are quite similar to the reliabilities for the original measures, calculated on the smaller group of school-going children. Third, we examined the correlations among the original and modified measures (Table 7).

Among the children in school, correlations of these modified measures with the original measures discussed above are quite high. For Whine, Depressed and Internalizing, original and modified measures are identical because the two items in question are not included in those composites. These findings suggest that researchers can use the modified scales when they wish to include preschool children without concern that omission of the two school items causes changes in measure reliabilities or in the relative scores of children attending school. Finally, we repeated the correlational and analysis of variance assessments of validity using the modified scales and the larger pool of children. The parallel correlations were quite similar (absolute discrepancies were generally less than .02; even the most discrepant pair of coefficients (.26 versus .20) were not significantly different). Finally, parallel ANOVA tables using the modified Total and Externalizing scales also documented the significant three-way-interaction depicted in Figure 1 and discussed above. This pattern of results suggests that the measures we have devised are robust with respect to omission of those items and inclusion of younger children in the analysis.

SUMMARY AND CONCLUSIONS

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 that also have demonstrated construct validity. 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 are inappropriate when sample sizes are small. 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. Indeed, in this research we have found several interactions, all consistent with expectations based on prior work.

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

- Achenbach, T. M. 1978. "The Child Behavior Profile: I. Boys aged 6-11." *Journal of Consulting and Clinical Psychology* 46: 478-488.
- Achenbach, T. M. and C. Edelbrock. 1978. "The Classification of Child Psychopathology: A Review and Analysis of Empirical Efforts." *Psychological Bulletin* 85: 1275-1301.
- Achenbach, T. M. and C. Edelbrock. 1983. *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, Vermont: Department of Psychiatry, University of Vermont.
- Bronfenbrenner, U., W. F. Alvarez and C. R. Henderson, Jr. 1984. "Working and Watching: Maternal Employment Status and Parents' Perceptions of Their Three-Year Old Children." *Child Development* 55: 1362-1378.
- Caspi, A., and G. H. Elder, Jr. 1987. "Childhood Precursors of the Life Course: Early Personality and Life Disorganization." Pp. in E. M. Hetherington, R. M. Lerner and M. Perlmuter (Eds.), *Child Development in Life-Span Perspective*.
- Gecas, Viktor. 1979. "The Influence of Social Class on Socialization." Pp. 365-404 in *Contemporary Theories about the Family, Volume 1*, edited by Wesley R. Burr, Reuben Hill, F. Ivan Nye, and Ira L. Reiss. New York: The Free Press.
- Hernandez, Donald. 1986. "Childhood in Sociodemographic Perspective." *Annual Review of Sociology* 12: 159-180.
- Hoffman, L. W. and F. I. Nye. 1974. *Working Mothers*. San Francisco: Jossey-Bass.
- Kim, Jae-On, and Charles W. Mueller. 1978. *Factor Analysis*. Beverly Hills, CA: Sage.
- Mechanic, David M. 1980. *Mental Health and Social Policy, Second Edition*. Englewood Cliffs, N.J.: Prentice-Hall.
- National Longitudinal Survey Handbook, 1987*. Columbus, Ohio: Center for Human Resource Research, The Ohio State University.
- Peters, Marie F., and G. C. Massey. 1983. "Mundane Extreme Environmental Stress in Family Stress Theories: The Case of Black Families in White America." *Marriage and Family Review* 6: 193-218.
- Peterson, J. L. and N. Zill. 1986. "Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children." *Journal of Marriage and the Family* 48: 295-307.
- Ruopp, R. and J. Travers. 1982. "Janus Faces Day Care: Perspectives on Quality and Cost." Pp. in E. Zigler and E. W. Gordon, (Eds.), *Day Care: Scientific and Social Policy Issues*. Boston: Auburn House.

Table 1. Cognitive and 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 Hatter) 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 children'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) Items vary by children's age; Subsets of items from the Infant, Preschool, and Elementary versions of the HOME scale; include mothers' reports and interviewer ratings.	All Children

^aOriginators of measures are indicated in parentheses.

Table 2. Factor-Based Behavior Problem Scales: Factor Structure and Item Statistics.

ITEM CONTENT	FACTOR-ITEM CORRELATION ^a					MEAN ^b	S.D.
	I	II	III	IV	V		
Disobeys at school	.64		.44			0.30	.51
Trouble getting along with teachers	.63					0.13	.38
Trouble getting along with other kids	.60	.49				0.28	.49
Bullies, cruel or mean to others	.53	.58				0.30	.50
Destroys own or another's things	.50					0.15	.40
Not liked by other children	.48					0.16	.42
Unhappy, sad or depressed	.43			.65		0.26	.47
Cheats or tells lies	.41	.40	.40			0.56	.57
Disobedient at home	.40	.57				0.59	.56
Stubborn, sullen, or irritable		.62	.43			0.62	.62
Strong temper, loses it easily		.61				0.55	.65
Argues too much		.58				0.82	.68
Sudden changes in mood or feeling		.45				0.80	.61
Has difficulty concentrating			.73			0.64	.68
Easily confused, seems to be in a fog			.60	.40		0.29	.52
Restless or overly active, cannot sit still			.57			0.72	.71
Impulsive, acts without thinking			.54			0.68	.58
High strung, tense, and nervous			.47			0.44	.62
Too fearful or anxious			.45	.40		0.46	.60
Difficulty getting mind off certain thoughts			.41			0.38	.54
Feels worthless or inferior				.62		0.23	.46
Withdrawn, doesn't get involved with others				.46		0.14	.38
Complains no one loves him/her				.40		0.41	.57
Demands a lot of attention					.70	0.67	.67
Is too dependent on others					.59	0.29	.52
Clings to adults					.51	0.46	.61
Cries too much					.46	0.25	.50
Does not seem to feel sorry after s/he misbehaves						0.49	.64
Eigenvalues	1.23	7.50	1.32	1.42	1.60		
% Variance	4.4	26.8	4.7	5.1	5.7		
Factor-Based Scales:							
Number of Items	9	8	10	6	4		
Cronbach's alpha	.80	.80	.80	.73	.64		

^a All factor- item correlations .40 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 1,792, due to listwise deletion of cases with any missing values. Factor labels are I, Peer Problems; II, Aggressive; III, Hyperactive; IV, Depressed; V, Whine.

^b Items were scored 0 (not true), 1 (sometimes true), or 2 (often true). Item statistics are reported for the 1,792 actual cases which had valid responses on all 28 items.

Table 3. Intercorrelations among First-Order Factors^a.

	Peer Problems	Aggressive	Hyperactive	Depressed	Whine
Peer Problems	---	.825	.730	.630	.390
Aggressive	.378	---	.740	.564	.403
Hyperactive	.403	.443	---	.685	.428
Depressed	-.364	-.239	-.266	---	.429
Whine	.332	.419	.443	-.287	---

Unweighted N=1,792.

^a The intercorrelations among the constructed factor-based composites are displayed above the diagonal; intercorrelations among the underlying factors, obtained in the factor solutions, are displayed below the diagonal. Signs differ for the underlying Depressed factor and the constructed factor-based scale because the underlying factor had negative loadings relative to the remaining factors, while the constructed scale was scored so that a higher value indicated more Depressed behavior.

Table 4. Higher Order Factor Analytic Solutions for Extracted Factors.

	<u>Forced Two Factor Solution</u>		<u>Default One-Factor Solution</u>
	<u>Externalizing</u>	<u>Internalizing</u>	<u>Total Behavioral Problems</u>
Peer Problems ^a	.567	.273	.626
Aggressive	.617	.229	.632
Hyperactive	.639	.267	.675
Depressed	-.191	-.574	-.452
Whine	.405	.494	.608
Eigenvalue	2.44	.803	2.44
% variance	48.8	16.6	48.8

^aInput for higher-order factor solutions is the correlation matrix among the underlying factors that was produced by the first-order factor solution (Kim and Mueller, 1978). Factor analyses used principal axis factoring. The default solution extracts as many factors as there are eigenvalues greater than 1.0; in this case, only one factor met the criterion. Two higher-order factors were estimated by over-riding the default criteria and specifying that two factors be extracted. No rotation is done for a single factor; varimax rotation was specified for the two-factor solution. Unweighted N=1,792.

Table 5. Evidence for Construct Validity: Bivariate Relationships with Factor-Based Scales.

FACTOR-BASED SCALES:	PEER PROBLEMS	AGGRESSIVE	HYPER- ACTIVE	EXTERNALIZING
<u>Current Economic and Occupational Conditions</u>				
Household Income	-.11 ^a	-.07	-.08	-.08
Mother's Annual Earnings	-.06	-.07	-.04ns	-.05
Poverty Status(1=yes)	.16	.10	.09	.12
Mother's Employment Status (1=employed)	-.10	-.09	-.07	-.09
Mother's Marital Status (1=married)	-.15	-.10	-.12	-.13
If Mother is Employed:				
Mother's Hourly Rate of Pay	-.05	-.06	-.05	-.05
Mother's Occupational Status	.03ns	-.01ns	-.01ns	-.01ns
<u>Family Background Characteristics</u>				
Mother's Educational Attainment	-.09	-.06	-.06	-.07
"Father's" Educational Attainment	-.10	-.07	-.08	-.08
Mother's Cognitive Achievement	-.15	-.07	-.13	-.13
Maternal Age at Child's Birth	-.07	-.05	-.04	-.06
Maternal Adult Shyness	-.01ns	-.01ns	-.03ns	-.03ns
Minority Status (1=non-white)	.06	-.04ns	.02ns	.01ns
<u>Child Characteristics</u>				
Age in Months	.05	.05	.02ns	.04
Gender (1=male)	.14	.06	.12	.11
Birth Order	.12	.10	.08	.10
Vocabulary Knowledge (PPVT)	-.19	-.09	-.16	-.16
Psychological Visits(1=yes)	.26	.22	.26	.28
Meds for Behavior Problems (1=yes)	.04ns	.06	.09	.06

Table 5. Evidence for Construct Validity: Bivariate Relationships with Factor-Based Scales (continued).

FACTOR-BASED SCALES:	DEPRESSED	WHINE	INTER- NALIZING	TOTAL PROBLEMS
<u>Current Economic and Occupational Conditions</u>				
Household Income	-.05	-.08	-.07	-.09
Mother's Annual Earnings	-.04ns	-.04ns	-.04	-.06
Poverty Status	.10	.05	.09	.12
Mother's Employment Status (1=employed)	-.08	-.04	-.08	-.11
Mother's Marital Status(1=married)	-.09	-.11	-.12	-.14
If Mother is Employed:				
Mother's Hourly Rate of Pay	-.04ns	-.02ns	-.04ns	-.04
Mother's Occupational Status	.01ns	.02ns	.03ns	.02ns
<u>Family Background Characteristics</u>				
Mother's Educational Attainment	-.04	-.04ns	-.04	-.06
"Father's" Educational Attainment	-.06	.02ns	-.02ns	-.04ns
Mother's Cognitive Achievement	-.11	-.14	-.14	-.14
Maternal Age at Child's Birth	-.06	.04	-.03ns	-.06
Maternal Adult Shyness	.09	.02ns	.04ns	.05
Minority Status (1=nonwhite)	-.00ns	.02ns	.01ns	.01ns
<u>Child Characteristics</u>				
Age in Months	.08	-.10	.00ns	.03ns
Gender (1=male)	.01ns	-.10	-.04	.08
Birth Order	.02ns	.09	.05	.07
Vocabulary Knowledge (PPVT)	-.13	-.11	-.13	-.16
Psychological Visits	.25	.16	.25	.29
Meds for Behavioral Problems	.04ns	.02ns	.04	.06

Note. Correlations are calculated with weighted data, using pairwise deletion for all cases with valid scores on all Behavioral Problem measures; maximum unweighted number of cases is 1,792. 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.

Correlations using "father"'s educational attainment are calculated only for children whose mothers are married and living with a spouse in 1986. The educational attainment of the mother's current husband is used, although this is not the child's biological father in all cases.

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 6. Multivariate Relationships with Total and Broad-Band BPI Scales:
Analysis of Variance.

	TOTAL		EXTERNALIZING		INTERNALIZING	
	F	P	F	P	F	P
<u>Main Effects:</u>						
AGE	.48	ns	1.26	ns	.02	ns
GENDER	10.36	.001	27.77	.000	3.59	ns
RACE	1.10	ns	1.16	ns	.93	ns
MOTHER'S MARITAL STATUS	16.95	.000	16.37	.000	14.44	.000
<u>Two-Way Interactions:</u>						
AGE by GENDER	3.32	.037	2.11	ns	4.88	.008
AGE by RACE	.22	ns	.34	ns	1.33	ns
AGE by M S	.39	ns	.28	ns	.55	ns
GENDER by RACE	.39	ns	.31	ns	.98	ns
GENDER by M S	.66	ns	.65	ns	1.67	ns
RACE by M S	.23	ns	.27	ns	.42	ns
<u>Three-Way Interactions</u>						
AGE by GENDER by RACE	.47	ns	.26	ns	1.55	ns
AGE by GENDER by M S	3.57	.007	3.39	.009	2.90	.02
AGE by RACE by M S	1.16	ns	1.03	ns	1.17	ns
GENDER by RACE by M S	.79	ns	1.21	ns	.64	ns
<u>Four-Way Interactions</u>						
AGE by GENDER by RACE by M S	1.37	ns	1.20	ns	1.21	ns
<u>EXPLAINED</u>	1.96	.000	2.18	.000	1.93	.000

Note. Age is categorized as: 4 and 5 year olds; 6, 7, and 8 year olds; and 9 years old and older. Gender is categorized as male and female. Race has three categories: Hispanic; Black; and Other. Mother's marital status is categorized as: never-married; currently married with spouse present; and formerly married. Analyses were conducted with the weighted sample of cases that had complete data for all variables; unweighted number of cases is 1792.

Table 7. Comparison of Complete and Modified Measures: Bivariate Correlations among School-Going Children and Reliabilities of Complete and Modified Measures.

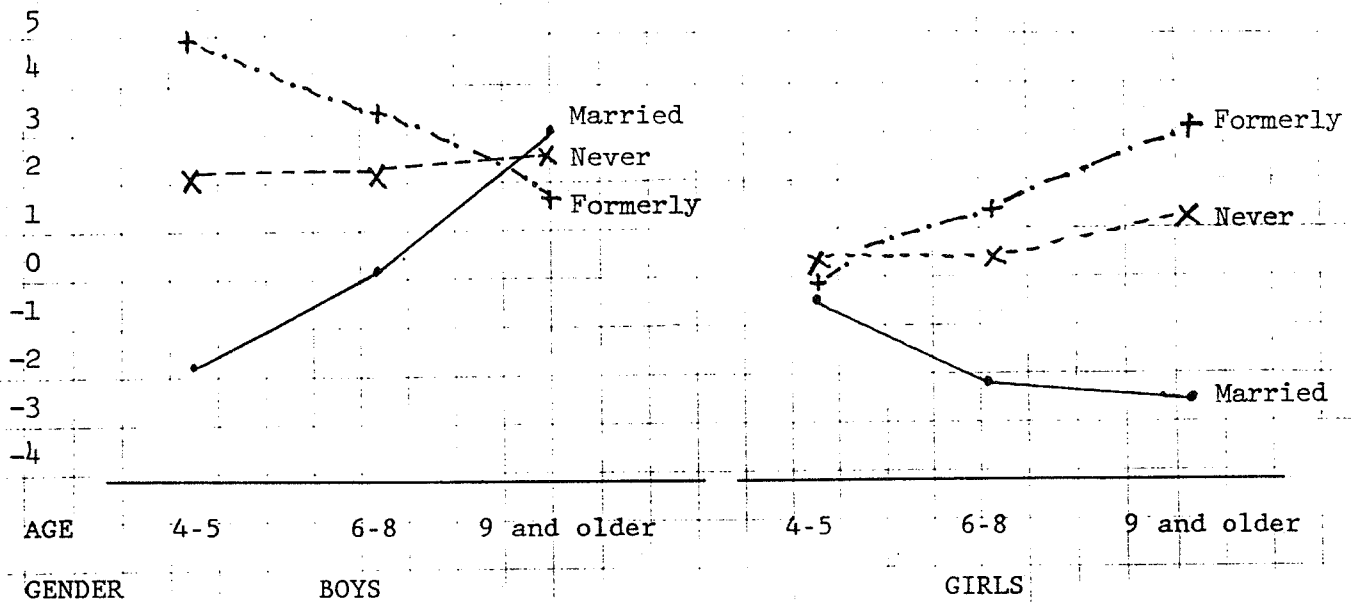
<u>Modified Factor-Based Composites^a</u>								
<u>Original Composites</u>								
	Peer Problems	Aggressive	Hyperactive	Depressed	Whine	EXT	INT	Total
Peer Problems	.968							
Aggressive		1.00						
Hyperactive			.989					
Depressed				1.00				
Whine					1.00			
Externalizing						.992		
Internalizing							1.00	
Total								.995
<u>Modified Measures:</u>								
Number of Items	7	8	9	6	4	18	10	26
Cronbach's Alpha	.73	.79	.78	.70	.66	.86	.75	.88
<u>Original Measures:</u>								
Number of Items	9	8	10	6	4	20	10	28
Cronbach's Alpha	.80	.80	.80	.73	.64	.88	.77	.90

^a Modified measures are identical to original measures for all scales that do not contain the two items asked only of school-going children; thus, correlations between original and modified measures are 1.0 for Aggressive, Depressed, Whine, and Internalizing.

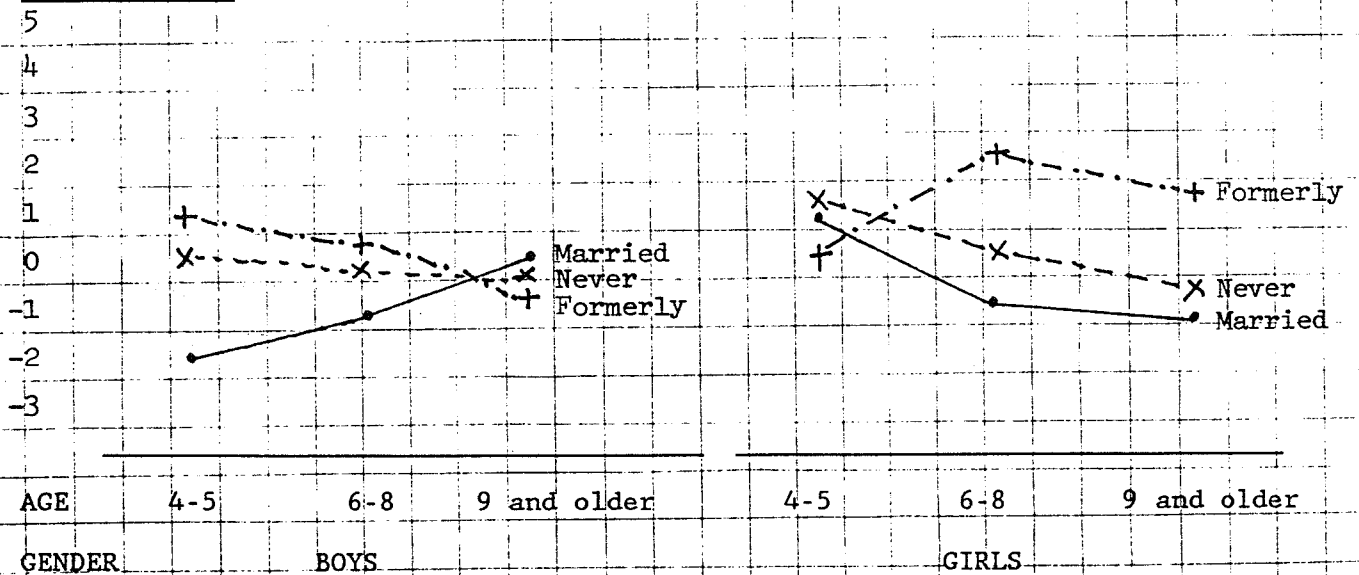
^b Reliabilities differ for the identical scales because those reported for the modified measures are calculated for the larger pool of both school-attenders and non-school attenders, while reliabilities for the original measures are calculated for the school-attenders only.

Figure 1. Externalizing and Internalizing Behavior Problem Scales:
Interactive Effects of Gender, Marital Status, and Child Age.

EXTERNALIZING



INTERNALIZING



Note. Age is categorized as: 4 and 5 year olds; 6, 7, and 8 year olds; and 9 years old and older. Gender is categorized as male and female. Race has three categories: Hispanic; Black; and Other. Mother's marital status is categorized as: never-married; currently married with spouse present; and formerly married. Analyses were conducted with the weighted sample of cases that had complete data for all variables; unweighted number of cases is 1792.