The NLSY79 Child Assessments

The NLSY79 data collection effort includes a substantial battery of assessment information about the children of female respondents who were interviewed in 1986 and biennially thereafter. The assessment information includes summary scores and, in some instances, subscores for all of the assessments. Where available, the file also includes national norms based on the raw scores. Tables 10 and 11 list the "raw" and normed scores available in the NLSY79 Child file. The data file also includes individual "raw" item responses for all the assessments administered in each survey round. Interviewer remarks associated with each assessment and "flag" items for several of the assessments indicating where prorations were necessary or where alternate scoring schemes were considered are also included. This issue will be addressed below in relation to specific assessments.

Members of the CHRR staff have examined the assessment data as carefully as possible while preparing the summary scores and the public use files. However, researchers who detect what appear to be significant data problems with the assessments are encouraged to contact:

Frank Mott   phone: (614) 442-7378   e-mail: mott.1@osu.edu
Paula Baker   phone: (614) 442-7375   e-mail: baker.21@osu.edu

Should a significant problem be detected, an effort would be made to immediately inform all data purchasers, to publicize the issue in the quarterly NLS newsletter, the NLSY79 Child website and to make corrections on all subsequent public releases.

The following section, organized on an assessment-by-assessment basis, includes information regarding how each assessment was scored, any relevant caveats, and where to find the variables relating to each assessment in the accompanying documentation. The discussion of the assessments generally follows the order in which they appear in the Child interview

schedules and in the documentation, beginning with the *Mother Supplement* assessments and followed by the *Child Supplement* assessments.

Additional Publications. Detailed information about the 1988 through 1992 Child data releases may be found in the *NLSY79 Children 1992: Description & Evaluation* and the *NLSY79 Child Handbook 1986-90 Revised Edition* (available at no charge from CHRR). For tabular material detailing the 1988 assessments see *Children of the NLSY79, 1988 Tabulations and Summary Discussion*. Tables describing the distributions of the 1994 through 1998 Child assessment scores can be found in *The NLSY79 1994 through 1998 Child Assessments: Selected Tables.* The material in the 1998 report is referenced in the assessment discussion that follows.

Who Was Assessed?

Some assessments are completed only once by a child--the first time he or she becomes age-eligible. Others are always completed by all age-eligible children. Finally, ten- and eleven-year olds complete almost all assessments for which they are age-eligible, regardless of whether or not they had previously completed the assessment(s). This provides users with an "index" group of children who ultimately will represent a large, more fully representative sample of early adolescent youth for analysis. More generally, beginning in 1994, children attaining age 15 (or older) in a given interview year were no longer given <u>any</u> of the assessments. Table 11 synthesizes the pattern of assessment administration across survey years.

NOTE: Users can rely on the child sampling weights to determine which children have assessment information in any given assessment year. Cases restricted to where a child sampling weight is greater than zero will yield a sample of assessed children in that year. However, these assessed children will not necessarily have a <u>valid score</u> on any particular

assessment in that year.   A series of flags (C00115.01-.06) indicates the interview and assessment status for 1998.

Changes in the Child Questionnaires

Two assessments, Memory for Location and Body Parts, were administered in 1986 and 1988, but have since been deleted from the data collection effort due to funding constraints. However, the 1986 and 1988 individual items and scores for these two assessments remain in the data file and are available to users. A brief description of these two assessments is included in this document. In 1992 and 1994, only the first two parts (A and B) of Verbal Memory were administered. This assessment has not been administered since 1994. In addition, beginning with 1996, the Self-Perception Profile is only administered to children age 12 and over.

In addition to the changes and additions to the *Child Self-Administered Supplement* discussed earlier in this document, the following changes have been made to the HOME sections in recent years. Beginning with the 1992 *Mother Supplement*: code categories were added to the questions on the relationship of the child's father/father-figure to the mother, and a category added to the parent efficacy question for older children. In addition, three questions on the following topics were added to Section 5 of the *Mother Supplement* in 1992: (1) rating of child's current school, (2) rating of the child's general well-being and prospects, and (3) degree of parent knowledge about child's friends. Most importantly, these questions are now completed by all mothers who have children of school age.

Beginning with 1994, one additional sequence has been added to the HOME assessment. Mothers are now asked how close each of her children feels to her, to the child's biological father, or to his/her stepparents (e.g., see Q16a for children under the age of three). In 1996, check questions asking if the child ever sees his or her father were dropped. Neither this change

nor any of those made in previous years affect the HOME score or subscores in any way. The components of all the scores have remained unchanged since 1986.

Assessment Completion Rates

Table 11 provides estimates of the number of children administered each assessment in 1998 and the completion rate of each assessment. As can be seen, the number of children undertaking and completing each assessment is quite substantial. Especially noteworthy is the number of black, white and Hispanic children available for separate racial/ethnic analysis. This latter factor is of particular importance for those assessments where there are major differences in outcome by race, or more importantly, where the linkages between critical explanatory inputs and assessment outcomes vary by race/ethnicity. It may also be noted from this table that, for the most part, the percent of children receiving valid scores is quite high, frequently over 90 percent, and racial variations in completion rates are generally quite modest.

Linking the Child Questionnaire to the Data File

Questionnaire items in the *Mother Supplement* are now uniquely identified in this document by the section and question number of the relevant item. Items in the *1998 Mother Supplement* are prefixed by an "MS." Thus, for example, question number 1 in section 1, Part A of the *1998 Mother Supplement* will be identified as MS981A01 because: (1) it appears in the *Mother Supplement* (MS); (2) it was administered in the 1998 survey round; and (3) it is located in "Section 1A: The HOME" and is question one. Through 1996, items in the Mother Supplement were identified by their relevant deck and column numbers. Thus, for example, in 1996, question 1 in section 1, part A was identified as MS960243 because (1) it appears in the Mother Supplement (MS), (2) it was administered in the 1996 survey round, and (3) it was

located in "deck 2" (see top of questionnaire page), column numbers 43-44 (see margin of questionnaire). Through 1992, items in the *Child Supplement* were similarly identified, except that the *Child Supplement* items were prefixed with "CS." Beginning with 1994, the items in the *Child Supplement* were administered using computer-assisted personal interview (CAPI) technology, so deck and column numbers are no longer available to identify items in that supplement. Items from the *Child Supplement* that are referenced in Tables 12 through 14 of this document are drawn from two sections at the end of the supplement entitled "Interviewer Evaluation of Testing Conditions" and "Interviewer Observations of the Home Environment." In these two sections, all items are identified by CS question name. For example, the item from Part A of the HOME Observations section of the *CAPI Child Supplement* asking the interviewer whether or not she saw the mother hugging the child is identified as CSOB-4A.

Description of the Child Assessments

In the following sections, each assessment is discussed in detail. Issues essential to using the 1998 version of each assessment are highlighted. Additional tabular material relating to each assessment appears in the *1998 NLSY79 Child Assessments: Selected Tables* mentioned earlier. A bibliography of references to research using the NLSY79 Child assessment data is available at

no charge from CHRR*: Child Assessment Research*, NLSY Maternal & Child Bibliography Series, No. 2 (CHRR, 2000).

The HOME-SF (Home Observation of the Environment - Short Form)

The Home Observation Measurement of the Environment-Short Form (HOME-SF) is the primary measure of the quality of a child's home environment included in the NLSY79 child survey. It has been extensively used as both an input in helping to explain other child characteristics or behaviors as well as an outcome in its own right—for researchers whose objective is to explain associations between the quality of a child's home environment and earlier familial and maternal traits and behaviors.

The HOME-SF is a modification of the HOME inventory (Caldwell and Bradley, 1984), a unique observational measure of the quality of the cognitive stimulation and emotional support provided by a child's family. The HOME-SF is about half as long as the HOME Inventory, an adaptation necessitated by survey time and cost constraints. More than half of the HOME-SF's items are multi-response maternal reports reworded from the original HOME Inventory's dichotomous observer ratings. The mother report items may be found in Section 1 of the *Mother Supplement* (see Appendix 1).

The HOME-SF is divided into four parts: the first for children under the age three (Part A); the second for children between the ages of three and five (Part B); the third for children ages six through nine (Part C); and the fourth for children ten and over (Part D). At several survey points, items included in the HOME-SF have been expanded. This is noted in the section above highlighting recent questionnaire changes. These items are not part of the actual computed HOME score or subscores.

Bettye Caldwell designed the Infant version of the original HOME Inventory and, with Robert Bradley, developed the Preschool and School Age versions. Bradley and Caldwell reviewed and approved the final draft of the Infant, Preschool, and Elementary HOME-SF versions used in the 1986 Mother and Child Supplements of the NLSY79, and Bradley was involved in a 1988 review.

Total HOME Scores. The total raw score for the HOME-SF is a simple summation of the recorded individual item scores and varies by age group, as the number of individual items varies according to the age of the child. The total HOME-SF score and the two subscores have one implied decimal place. For example, a score of 30 is really 3, and so on. In addition, total scores were imputed for children where one or more of the component items had inadvertently been left unanswered. The imputation procedure assigns an average value, derived from all those items that had been completed, to each of the unanswered items. Proration flag variables specify the number of items which require imputation for the different age groups; a score of zero on this proration flag variable means that all individual component items were answered. For the two subscores specified below, a more stringent proration rule was followed: scores were derived only for cases where no more than one item was missing.

Cognitive Stimulation & Emotional Support. In addition to the overall HOME-SF score, the Child file includes two subscores: a cognitive stimulation and emotional support score. The (questionnaire item) components of the total as well as cognitive stimulation and emotional support subscales are specified in the HOME appendix to this document. Because there are no appropriate national norms available for the overall HOME-SF score or its components, we provide internally normed standard and percentile scores for the overall HOME-SF scores as well as for the cognitive stimulation and emotional support subscores.

Norms. The internal norms were developed using standard normal curve assumptions. Children were normed on a single year of age basis with each (weighted) single year age of age group being assigned a standard score mean of 100 and standard deviation of 15. Percentile scores were derived from the standard scores using an inverse normal routine. To the extent that the single year of age data deviate from normality, this procedure produces less than optimal results. An alternate percentile score can be generated using the empirical cumulative distribution function by age computed using the sampling weights. That frequency could be used to crosswalk from raw score to percentile score.

Recoding of HOME Items. Prior to constructing an overall score as well as the two subscores for the HOME-SF, all of the individual items were translated into dichotomous zero-one variables and then appropriately summed. The precise recoding used in computing the HOME scores can be found in Appendices A1 and A2. The dichotomous HOME items for 1998 appear on the CD in the ASSESS98 area of interest, reference numbers C18996.-C19084.

Recoding Discipline Items. Several of the HOME-SF items required extensive initial recoding in order to fully utilize the verbatim responses originally coded as "other." The HOME-SF Part B contains items (Mother Supplement question MS981B08) concerning mother's response to the child hitting her. Parts C and D of the HOME-SF contain items (question (MS981C22 and MS981D21 respectively) concerning the mother's response to the child swearing at her. Both items are coded "1" if the parents' response is moderate, defined as without harsh reprisal.

The Part B item contributes to the HOME-SF scale scores only if certain alternatives ("send to room," "talk," "ignore," and "give a chore") are selected and the "other" alternative is without harsh reprisal—that is, if a mild reaction is the first response. The Part B item is scored zero of any of the following are selected: "hit," "spank," or the "other" alternative is harsh.

Harshness is arbitrarily defined as either extensive or excessive deprivation (time-out longer than two hours; deprivation longer than two days) or physical punishment (firmly grasping the child, spanking then talking, or talking then spanking).

The Part C item was scored similarly. Yelling back and withdrawal of love, while perhaps emotionally harsh, were scored as mild (score of 1) because they are not physical responses. The item is scored zero if "spanking" is selected or if the "other" alternative is excessive (longer than three hours of time-out; longer than three days of deprivation) or if physical means ("eat soap") are the first types of punishment selected. Examples of verbatim scores as harsh are "break him up," "spank and ground for two weeks," and "spank then explain why." If the length of time was not specified ("send to room") then it was assumed to be a moderate amount of time, scoring the item was mild. Other examples of verbatims scored mild are "never happens," "depends on the situation," "stand in corner until apologizes." A classic mild response (conveying no discipline) was "give him something to eat." A few other verbatims should be noted. One respondent with three children "hit" and commented, "Then say I'm sorry and laugh." Another mother of two checked "hit" saying, "But not like I'd hit an adult."

The HOME Scores. As indicated above, the items completed by children are dependent on their age; children under age 3 years, 3 through 5, 6 through 9, and ten and over follow different question sequences. The actual items and the recoding instruction may be found in the Appendix A1 and Table 15. The reference numbers for the children's HOME scores in 1986 through 1994 may be found in Table 8. Table 10 lists the references numbers assigned to the 1996 and 1998 HOME scores. In addition to the raw scores (which have different reference or "C" numbers depending on the HOME section the child completed), the reference numbers for the internally normed standard and percentile scores may also be found in that table. All

children under the age of 15 living with their mothers in 1998 were eligible for the HOME assessment (until 1994, all children, regardless of age, had a HOME supplement completed by their mother). Thus, children born by the 1986 survey date (and still below the age of 15 in 1998) may have seven rounds of HOME scores available. Children born between 1986 and 1988 may have six HOME scores and children born since the 1988 survey can have up to five HOME scores—assuming of course that their mothers completed a HOME assessment for them at the relevant survey points. Note also that whereas the raw scores are specific to a child's age at a particular survey point, a single set of normed scores is created for each survey round, regardless of the child's age.

HOME Completion Rates. Table 11 indicates the number of children at different ages whose mother or guardian completed a HOME supplement in 1998 and presents HOME completion rates by age, race, and ethnicity. Overall, 92.4 percent of children under age 15 have a completed HOME assessment. Completion rates are much lower for the youngest children, because the interviewer observation items at the end of the *Child Supplement* are less likely to be completed for this age group. Since there are no assessments directly addressed to children under the age of four, interviewers are less likely to see the mother in direct contact with the child, and thus are less able to meaningfully answer the items that require direct observation of mother-child contact. This is even a more significant issue with respect to the emotional support subscore. This is partly because the conditions permitting proration of scores are more stringent, as noted above.

Validity & Reliability. The HOME assessment is among the most often used of all the assessments. It is widely employed both as an input, predicting many other child outcomes, and as an outcome in its own right. As mentioned earlier, the *NLSY79 Child Assessment Research Bibliography*, available from the Center, includes an extensive section listing research that has

used the NLSY79 HOME data. The *NLSY79 1990 Child Handbook* describes this research in

detail, emphasizing work that has examined the reliability and validity of the HOME. *The NLSY*

Children 1992 provides further evidence regarding linkages between the HOME scale and

subscales, and a variety of family and maternal antecedents. Finally, Tables 1.1 through 1.8 in

The 1998 NLSY79 Child Assessments: Selected Tables provide detailed distributions by age and

race/ethnicity for the overall scores as well as the two subscores.

Temperament (How My Infant Usually Acts)

At the time of the 1986 NLSY79 Child survey design, no single instrument seemed

adequate to use for measuring child temperament, within the context of a large national survey

administered by lay personnel. As a result, a Temperament scale was developed, based on

Rothbart's Infant Behavior Questionnaire, Campos and Kagan's compliance scale, and other

items from Campos.

Because the child's temperament is partially a parental perception, the behavioral style of

children in the NLSY79 was measured by a set of maternal-report items (for all children younger

than seven years) and interviewer ratings (in 1996, for children three years or older). The

maternal scale "How My Infant Usually Acts" addresses the activity, predictability, fearfulness,

positive affect, and friendliness of infants below age one. "How My Toddler Usually Acts"

addresses the fearfulness, positive affect, and friendliness of one-year-olds. "How My Child

Usually Acts measures the compliance and attachment of two- and three-year-olds and

additionally, the friendliness of children aged four through six. For children ages three through

six, the interviewer rates the child's shyness when first introduced, shyness at the end of the

session, and the child's cooperation, interest and motivation, energy, persistence, and attitude

toward and rapport with the interviewer during the assessment. All of the scales were available in English and Spanish.

The Temperament Scores. A total of ten distinct scores tap various dimensions of temperament, but not all dimensions are appropriate for all ages. The specific (questionnaire) item components of each score, as well as its age appropriateness, are indicated in Appendix B. The complete listing of available assessment scores for 1986 through 1998 may be found in Tables 8 and 9.

The behavioral tendencies of the children are rated by the mother on a five-point scale, ranging from Almost Never (value of 1) to Almost Always (value of 5). The scores of the various scales are computed by simply summing the individual items in the scale. Some items are recoded in reverse where appropriate before summing and these items are indicated by an (R) next to the individual item in Table 13. If any item component of a subscale was missing, that subscore was not computed. Since no appropriate national norms are available for this assessment, only raw scores are provided.

Changes in Scoring. An important change, of necessity, was made beginning with the 1990 Temperament scoring. Because children under the age of four in some survey rounds are not administered any of the *Child Supplement* items, it is necessary to truncate two scale addressed to younger children, the difficulty composite score for children between the ages of 8 months and 23 months and the friendliness scale for children under the age two. For researchers requiring comparability over time, abbreviated and unabbreviated versions of the scores for 1986 and 1988 are included in the public use file.

Completion, Validity, & Reliability. Those considering the use of the temperament scores may wish to examine the detailed Temperament tabulations available in *The 1998 NLSY79 Child Assessments: Selected Tables* (CHRR, 1998b) as well as the evaluation of the

temperament reliability and validity included in *The NLSY Children 1992* (Mott, et al., 1995).

The latter document examines, within a multivariate context, the extent to which selected

temperament scores are independently linked with a wide range of demographic and

socioeconomic antecedents as well as being predictive of other child assessments in subsequent

survey rounds. In general, completion rates for this assessment are quite high, often being above

90 percent, as may be noted in Table 2.1 to 2.12 in *The 1998 NLSY79 Child Assessments:*

Selected Tables (CHRR, 1999a)*.*

Motor and Social Development

        The Motor and Social Development scale (MSD) was developed at the National Center

for Health Statistics to measure dimensions of the motor, social and cognitive development of

young children from birth through three years. The items were derived from standard measures

of child development (the Bayley, Gesell, Denver), which have high reliability and validity (Poe,

1986).   Analyses by Child Trends of these same items in the 1981 *Child Health Supplement to*

the National Health Interview Survey (a large national health survey of 2,714 children up to age

four) established the age ranges at which each item's developmental milestone is generally

reached by U.S. children (Peterson and Moore, 1987).   Based on the child's age, NLSY79

mothers answer fifteen age-appropriate items out of 48 motor and social development items.

These items have been used with a full spectrum of minority children with no apparent difficulty.

A Spanish version of the schedule is available to mothers whose principal language is Spanish.

        Scoring Motor & Social Development. The NLSY79 Motor and Social Development

assessment has eight components (parts A through H), which a mother completes contingent on

the child's ages. Part A is appropriate for infants during the first four months of life (i.e., zero

through three months) and the most advanced section, Part H, is addressed to children between

twenty-two and forty-seven months. All of the items are dichotomous (scored either zero or one) and the total raw score for children of a particular age is obtained by a simple summation (with a range 0 to 15) of the affirmative responses in the age appropriate section. Associated with each raw score is a series of norms: an overall percentile and standard score as well as same-gender age appropriate percentile and normed scores. That is, boys were given male national norm scores and girls were given female national norm scores, in addition to both genders receiving the combined gender norms. All these normed scores were constructed by CHRR using data from the nationally representative sample in the 1981 Child Supplement to the National Health Interview Survey (National Center for Health Statistics, 1984). The reference numbers for the various raw scores and overall and same gender normed scores for Motor & Social Development may be found in Table 9 and Table 10.

Norms. The norms are grouped into fairly narrow age categories reflecting the extreme sensitivity of a child's level of development to his or her age: following a (four month) zero through three months age break, the four through thirty month age range was normed by successive three month age groups with the thirty-one through forty-two month range being normed according to three successive four month categories, followed by one five month (forty-three through forty-seven month) category. No proration was attempted on this assessment since the proportion of missing items is modest and there was some question about the appropriateness of the procedure, given that later items in the assessment tend to be more difficult than earlier items, and non-response is not random across items.

Caution should be exercised when interpreting results for three-year-olds, the oldest group completing this assessment. The Motor and Social Development assessment tends to "top out" for three-year-olds and does not provide a sensitive ceiling for these older children. *For this reason, researchers using the assessment should include an age control in any multivariate*

analyses even when they are using normed scores. In general, the distribution of scores for NLSY79 children on this assessment tends to be more peaked for the youngest and oldest children (e.g., see CHRR, 1999; Table 3.3).

While not described in these tables, it is also useful to note the reported gender differences at the youngest ages. Infant girls score significantly higher than their male counterparts, consistent with other evidence regarding early gender differences in motor and social development. Researchers interested in analyzing boys or girls separately are reminded that separate gender-specific norms are available.

Completion, Validity, & Reliability. As may be seen in Table 12, the overall completion rate for MSD in 1998 is about 87 percent, with a lower completion rate being evidenced for Hispanic and black children. This overall completion rate is down several percentage points from recent rounds. A substantial proportion of the non-completions resulted from situations where no attempt was made by the mother to even begin assessment. Readers wishing more detailed information about the validity and reliability of this assessment may wish to examine the discussions of MSD in the *NLSY79 Child Handbook* (Baker et al., 1993) and *The NLSY Children 1992* (Mott, et al., 1995). Additionally, an appendix to the *NLSY79 Child Handbook* (Mott, et al., 1995) includes the complete norming tables for this assessment.

The Behavior Problems Index

The Behavior Problems Index was created by Nicholas Zill and James Peterson to measure the frequency, range, and type of childhood behavior problems for children age four and over (Peterson and Zill, 1986). Many items were derived from the Achenbach Behavior Problems Checklist (Achenbach and Edelbrock, 1981) and other child behavior scales (Graham and Rutter, 1968; Rutter, Tizard and Whitmore, 1970; Kellam et al., 1975).

Parent respondents to the 1981 Child Health Supplement of the National Health Interview Survey were asked an extensive series of structured questions concerning the child's problem behaviors and use of mental health services (NCHS, 1982, pp. 100-102). The specific questions asked varied somewhat depending on the age of the child. The behavior problem items utilized in the NLSY79 were developed from these items.

Scoring the BPI. The Behavior Problems total score is based on responses from the mothers to 28 questions in the *Mother Supplement* (items 1-26, 31, and 32 in the 1998 Behavior Problems scale) that ask about specific behaviors that children age four and over may have exhibited in the previous three months. Items 27 through 30 in the Mother Supplement BPI section are <u>not</u> part of the Behavior Problems scale. They were added subsequently by CHRR staff to tap dimensions that are particularly relevant for older children. Three response categories ("often true," "sometimes true," and "not true") were used in the questionnaire.

Scoring. For the overall Behavior Problems scale as well as the set of six subscales defined below, responses to the individual items are dichotomized and summed to produce an index for each child. In this recoding process, each item answered "often" or "sometimes true" is given a score of one. Each item answered "not true" is given a score of zero. Thus, higher scores represent a greater level of behavior problems. Two of the items (Q.31 and Q.32 in the Behavior Problems sequence) are appropriate only for children who have ever attended school. <u>Only the overall score and the antisocial subscore use these two items</u>. Thus, for these assessments, parallel raw scores are computed for children in school and children not yet in school.

The BPI Scores. Factor analysis was used to determine the six subscores alluded to above according to the following domains: (1) antisocial behavior, (2) anxiousness/depression, (3) headstrongness, (4) hyperactivity, (5) immature (6) dependency and (7) peer conflict/social

withdrawal. Appendix D1 of this document displays the components of these subscales. The

procedures used to define these subscores are detailed in the *NLSY79 Child Handbook* (Baker et

al., 1993).

Externalizing/Internalizing Scales.   CHRR has also prepared an alternate revised

trichotomous BPI scale as well as two subscales measuring a child's tendency to internalize or

externalize behaviors. These three scales, preferred by some users, are constructed from items

that are <u>not</u> dichotomized but are recoded from 1 (sometimes), 2 (often), 3 (never) to 0, 1, 2 with

the following recoding: $3 = 0$, $2 = 1$, $1 = 2$ before summing. The exact composition of these

externalizing and internalizing scales can be found in Appendix D1.

Norms. All of the above scores and subscores are available for all age eligible children

who were assessed biennially between 1986 and 1998. For all of the above except the last three

(the non-dichotomous external, internal and total scores which were not recoded), overall as well

as "same-gender" normed scores have been created based on data from the 1981 National Health

Interview Survey. (Girls are systematically likely to evidence "better" behavior on most of these

scales.) These normed scores include both percentile and standard scores (with a national mean

of 100 and a standard deviation of 15) and are based on single year of age data. For children

below the age of six, separate norms are created for children in and out of school.

Nationally normed percentile and standard scores are provided for the three trichotomous

scores, but normed "same-gender" scores are not available. The overall behavior problems score

based on non-recoded items is identified by the word "revised" in the variable title. With regard

to the six subscores, the user is cautioned that the range of normed outcomes is quite constrained,

because of the limited number of possible responses for some of the subscores. As with the other

Mother Supplement assessments, a user who wishes to select a sample of children of a particular

age should access the *Mother Supplement* child age variable. Users will find the reference

numbers for all of the Behavior Problems scores in Table 9 and Table 10. Additionally, the components for the various scores and subscores may be noted in Table 14 and Appendix D1. Note: Normed scores are not available for the Dependency subscale for children aged 12 and over. The BPI norming tables may be found in Appendix D2.

Imputation. Since 1992, imputed values have been assigned for the overall dichotomous raw score for all children for whom one item was missing. Norms are, of course, also provided for all those children. Similarly, beginning in 1994, scores have been imputed for the overall internal and external subscales where only one item is missing. The overall trichotomous score is <u>not</u> imputed as of 1998. The overall dichotomous raw score includes one extra digit, with the final digit representing one implied decimal place. The external and internal raw scores have been rounded to the nearest full digit. Imputation flag variables are available that identify those cases that have been prorated. In no instance, does imputation involve very many cases.

Validity & Reliability. The Behavior Problems Index is among the most frequently used of the NLSY79 child assessments, both as an outcome in its own right and as a robust predictor of a wide range of child attitudes and behaviors. The *NLSY79 Child Handbook* (Baker et al., 1993) discusses the available literature on this assessment. *The NLSY Children 1992* includes a discussion of the reliability and validity associated with this assessment (Mott, et al., 1995). The most recent bibliography of NLSY79 research using the Child assessment data cites research based on these scales (CHRR, 2000).

The overall completion rate for the Behavior Problems scale is about 93 percent with Hispanic children having somewhat lower levels of completion. This racial/ethnic variation is not as pronounced as in earlier survey rounds (Mott, 1998).

It is important to note that, although a fully representative sample of children would be expected to have a mean standard score of 100, the mean for the NLSY79 Child sample is 103.5

(Table 4.3 in CHRR, 1999a). In previous evaluations (Mott 1998), we had speculated that this reflected the fact that the NLSY79 children are not yet fully representative of a national cross-section of American children and somewhat over-represent children born to younger and less-educated women. This becomes less of an issue with the passing of time as the cohort of mothers approaches the end of their childbearing years. For example, whereas the NLSY79 children over the age of ten in 1996 have mean behavior problem scores of 106, the mean score for children aged 4 or 5 is about 98! While the age pattern appears somewhat erratic over time, there is indeed evidence suggesting that the sample of NLSY79 children may well ultimately have normed scores not substantially different from the overall 1982 norming sample. The 1998 Assessment tables report includes distributions for all of the 1998 Behavior Problems Scales (see Tables 4.1-4.17 in CHRR, 1999a.).

Parts of the Body (1986 and 1988 only)

The Body Parts assessment was completed by age-eligible NLSY79 children in 1986 and 1988 only. Developed by Jerome Kagan of Harvard University, Parts of the Body attempts to measure a one- or two-year-old child's receptive vocabulary knowledge of orally presented words as a means of estimating verbal intellectual development. The interviewer names each of ten body parts and asks the child to point to that part of his or her body.

Scoring Body Parts. The child's score is computed by summing the items that a child correctly identifies (C7972. for 1988 and C5799. for 1986). Thus, a minimum score is 0 and a maximum score is 10. No proration was attempted since the later items are more difficult than the earlier items in the sequence. A Spanish version of this assessment was available for use with young Hispanic children. A complete protocol for this assessment may be found in the *1988 Child Supplement.*

Because there was some inconsistency in the way the interviewers interpreted the instructions, the Body parts assessment was scored in 1988 using three alternate criteria. First a child had to answer each of the ten items either correctly (1) or incorrectly (2) on at least one of the two attempts (see page CS-18 in the *1988 Child Supplement*). If scoring was completed according to this criteria, then the case was coded a "1" on the Body Parts scoring criteria flag (C7973.). A second, less restrictive criterion, allowed some of the individual items to coded "3" (no answer) on some of the attempts. For this subset of children, a code of 3 was treated as an incorrect response and the overall assessment scored accordingly. These cases can be identified by a value of 2 on the Body Parts criteria flag. Children for whom virtually all the responses were coded 3 (and translated into incorrect responses), received a value of 3 on the Body Parts criteria flag. Thus, users may restrict analyses to the more constrained sample or opt to include only children who had been scored according to the less conservative definitions. As with all the assessments in the *Child Supplement*, anyone who plans to extensively use a particular assessment is strongly urged to evaluate the scoring schema and data quality according to their own criteria. While we have made every effort to create scores that are faithful to the intentions of the assessment designers, there are instances where researchers could reasonably disagree about what precise scoring procedures should be utilized. Especially relevant to this issue is the fact that this assessment was given to very young children for whom there could be considerable ambiguity in differentiating between "incorrect" and "non" responses.

Norms. As no appropriate national norms are available for scoring this assessment, we provided (for 1988) internally normed standard and percentile scores (see Table 9). No normed results are provided for 1986. As the raw score on this assessment is extremely sensitive to the age of the child, users of the raw scores are encouraged to apply appropriate techniques that permit analytical comparisons of children across different ages. When controlling for age, the

user should select the appropriate *Child Supplement* age variable that specifies the child's age (in months) as of the *Child Supplement* interview date.

Completion, **Validity & Reliability**.   Notwithstanding the availability of a Spanish version of this assessment in the NLSY79, the user should proceed cautiously when interpreting its reliability and validity, particularly with regard to minority and relatively more disadvantaged children. It appears that a child's score may be quite sensitive to the child's English language capabilities as well as rapport with the interviewer. In 1986, the non-completion rate for this assessment was about 17 percent. For about half of the completed assessments, a child is reported to have not responded on at least one question, requiring the alternate assumptions with regard to scoring we describe above to be made. For a more complete discussion of the reliability and internal validity of this assessment and the Memory for Location assessment, please see pages 30-31 in *The NLSY Children 1992* (Mott et al., 1995).

Memory for Locations (1986 and 1988 only)

The Memory for Locations assessment was completed by age-eligible NLSY79 children in 1986 and 1988 only. It was developed as a measure of a child's short term memory and has been extensively used by Jerome Kagan of Harvard University (Kagan, 1981). The child, aged eight months through three years, watches as a figure is placed under one of two to six cups. The cups are screened from a child's view for one to fifteen seconds; the child is then asked to find the location of the figure. Items increase in difficulty as the number of cups and/or the length of time during which the cups are hidden from view increases. A child's score is based on his or her ability to select the cup hiding the figure.

Scoring Memory for Locations.   The number of individual items that a child can potentially answer in this assessment is contingent on the age of the child. Children between the

ages of 8 and 23 months start with item 1, the easiest question; children who are at least two years of age begin with item 4, and children age three start with item 7. A child's score is based on the highest (most difficult) question answered. A child who cannot answer the entry item receives a raw score of zero regardless of where he or she enters. Otherwise, if Q.1 is the highest item answered correctly, the child receives a score of 1. The maximum score is 10, if the tenth or final item is answered correctly. A child under two years of age is eligible to receive a score between zero and ten; a child age three, by virtue of the fact that he/she enters at item seven, can only receive a raw score of 0,7,8,9,10. Because external norms were not available, internally normed standard and percentile scores were developed. These 1986 and 1988 variables are identified in Table 9. The user is still advised to use to normed scores cautiously because of the unusual distribution of raw scores described above.

Because of the complexity of administering this assessment, a number of responses were not coded precisely according to the theoretical decision rules. On the advice of the assessment developer, children who followed a sequence that might have led to "extra learning" (as part of the assessment administration process) were still scored. For example, if a child was asked Q.1B after having correctly answered Q.1A, the child was scored and not given an "invalid skip" code, even though, theoretically, the child was supposed to proceed directly from Q.1A to Q.2A. In addition, a careful examination of the individual responses suggests that there were a number of children who began the assessment at an improper entry point but who ended up at a level where they would, in all likelihood, have wound up anyway. In these instances, a score was provided for the children and these cases were "flagged" with a code of "2" on the Memory for Location flag variable (C7977. for 1988 and C5782. for 1986). A code of "1" on this flag includes all scored cases except those defined as 2's. Researchers who plan to use this assessment extensively should carefully examine the actual response patterns to individual items. Individual

researchers may choose to impose scoring criteria that are more or less stringent than those used in computing the raw scores provided in this data file.

It is important to note that this assessment displays a clear tendency to "top out" for the oldest children in the sample. That is, a very large proportion (63 percent in 1986) of all three year olds and 32 percent of two year olds received the maximum raw score of ten on the assessment. A relatively normal distribution may be in evidence only for children below the age of two.

The Body Parts and Memory for Location assessments were deleted from the NLSY79 child assessment package following the 1988 Child data collection effort, partly because of funding constraints and partly because of the greater difficulty in administering them to children using lay interviewers in a home environment. For example, it was quite difficult to make an unambiguous determination as to whether a child was unable to respond or whether he/she was just shy. Additionally, it was sometimes difficult to be definitive regarding the direction in which a child was pointing, either toward a cup or toward a body part.

Finally, early evaluation of these two assessments in 1986 showed little in the way of significant linkages between a wide range of socio-economic antecedents and these two outcomes. However, some recently completed evaluative research suggests that these two assessments may be useful independent predictors of cognitive development (Mott, et al., 1995) since Body Parts and Memory for Location scores in 1986 are highly significant predictors of Peabody assessments in 1992. Thus, in standard multivariate analyses, it appears that these early child cognitive measures are indeed useful predictors of standard aptitude and achievement measures six years later.

McCarthy Scales of Children's Abilities - Verbal Memory (1986-1994)

The Verbal Memory subscale of the McCarthy Scales of Children's Abilities, last administered in 1994, assesses a child's short-term memory in response to auditory stimuli. The Verbal Memory subtest selected for use in the NLSY79 Child is only one of six scales that form the complete McCarthy assessment battery. Verbal Memory was administered by first asking the child, age three through six years, to repeat words or sentences said by the interviewer (Parts A and B). Then the child listens to and retells the essential aspects of a short story read aloud by the interviewer (Part C).

From 1986 to 1990, both the word and sentence components as well as the story part of the assessment were administered. In 1992 and 1994, administration was limited to the word/sentence component of the assessment. (After 1994, for cost reasons as well as for a number of quality related reasons discussed below, administration of this assessment was discontinued.)

Verbal Memory has typically been completed by children between the ages of three and six; although in 1990, administration was limited to ages four to six. Additionally, in all administration years, it was only administered to age-eligible children who had not previously (in a prior round) completed the assessment. The precise administration pattern may be noted in Table 11.

Scoring Verbal Memory. In the first half of the word-sentence component of the assessment (Part A), the score which the child received was contingent on the child repeating a series of words, ideally in the same sequence as they were uttered by the interviewer. In Part B of this first section, the child was scored according to the number of key words that he or she repeated from a sentence read by the interviewer. The combined total score for Parts A and B determined whether the story (Part C) was administered. In Part C, the child was read a story

paragraph and then scored on the basis of his or her ability to recall key ideas from that story. National norms are available for this assessment, so children were assigned normed scores based on his or her performance in comparison with a nationally representative sample.

The number of correct responses to the words and sentences on pages 50 and 51 in the 1994 *Child Supplement* (the last year the assessment was administered) were combined to generate one total raw score. Appropriate national norms are available in the McCarthy manual (McCarthy, 1972: 205). Thus, percentile and standard scores are available for linking with the raw scores. The specific identification for these raw and normed scores for 1986 to 1994 may be found in Table 9.

As noted, the 1986 through 1990 rounds of data collection, the Verbal Memory assessment included a "Part C" or a "Story" section. Children who received this assessment in 1986-1990 received two scores in each year. Entry into the "Story" was contingent upon receiving a minimum combined score of 8 on Part A plus Part B. The researcher may note that there were a few instances of children entering and receiving a score on Part C who had received an invalid skip score on Part A and Part B. While it may not have been possible to score A and B for various reasons, the available information was sufficient for the scorer to be confident that the A and B score was at least 8. Children who received a valid score of less than 8 on Part A and Part b were automatically assigned a zero on Part C. This explains the considerable heaping at the zero outcome for Part C.

The scoring on Part C is a simple summation of the number of key words/phrases identified correctly from the paragraph on page CS-36 of the 1990 *Child Supplement*. No proration was attempted for missing responses. The individual items may be seen on page CS-38 of that supplement. A total raw score and two normed scores were generated for Part C in 1986 through 1990 (Table 9).

From an analytical perspective, the prospective user may note that the distribution of the percentile and standard scores for Part C are somewhat uneven, reflecting the fact that the Part C outcome allowed for only 12 possible responses (0 and 1 through 11) with a major heaping as noted, at the zero category. The fact that the percentile/standard scores assigned to the various raw scores varies by the age of the child helps to smooth the normed response somewhat. However, the user is encouraged to examine the pattern of normed responses before proceeding with his or her research. As with all of the assessments in the *Child Supplement*, the <u>Child Supplement age</u> variable should be used when stratifying the sample by age of child.

Validity. While this subscale has a high face validity regarding what it purports to measure, the user should be sensitive to the fact that the scoring of Part C, the story section, undoubtedly includes an element of subjectivity. Interviewers can, in some instances, disagree regarding whether or not a child's specific response was indeed a "correct" or "incorrect" interpretation of an aspect of the story. Also, to some extent, the verbatim verbal responses recorded by the interviewer could in some instances be coded in different manners by different interviewers. In order to test this latter premise, NORC had the 1986 verbatim responses for about 400 children coded independently by two coders. There was complete agreement between coders for 92 percent of the respondents.

At a different level, there is also some possibility that the Part A response patterns reflect a lack of precision in the instruction—an ambiguity that also exists in the McCarthy manual. The instructions (for Part A) only ask the child to repeat the words which the interviewer reads to him or her, but does not specify that the words should be repeated in the same sequence. However, in the scoring, the respondent loses a point if the words are repeated out of sequence. Thus, the extent to which the words were repeated in or out of sequence may have been a

function of how the instructions were understood, an artifact that could attenuate the reliability of the Part A score.

Completion Rates. The 1994 completion rate for Parts A and B was only about 82 percent, below the completion rate for all of the other child-administered assessments. Hispanic children had a completion rate of only 77 percent, substantially below that for other children. Thus, as with some of the other assessments, there is surface evidence that language constraints come into play when evaluating the reliability and potential validity of this assessment. With regard to this assessment, it is important to note that a Spanish translation was _not_ utilized. This test measures *English* language verbal retention. Thus, a language bias is clearly possible and implied for at least some children. Hispanic children and children of less educated mothers are heavily over-represented among those who could not be scored—the "invalid response" subset.

Verbal Memory has been one of the most difficult of the assessments to administer because of the ambiguity involved in determining whether a child does not know an answer or is just shy (see Baker and Mott, 1995, for a discussion of this issue and its impact on the assessment). This is primarily an issue with younger children who had not previously been tested or had not been in a formal school environment. With the introduction of the CAPI administration procedures in 1994, one additional problem became apparent. The number of cases scored "zero" increased substantially, but interviewer comments suggest that many of these cases really should have been "non-completions." This is discussed in detail in Baker and Mott (1995). For the reasons noted above, this assessment should be used cautiously. Additional discussion relating to the reliability and validity of this assessment as well as how it has been used by other researchers may be found in the *NLSY79 Child Handbook*   (Baker, et al., 1993) and in *The NLSY Children 1992* (Mott, et al. 1995).

Self-Perception Profile for Children

The Self-Perception Profile for Children (SPPC) is a self-report magnitude estimation scale that measures a child's sense of general self-worth and self-competence in the domain of academic skills (Harter, 1982, 1985). Harter's instrument taps five specific domains of self-concept as well as global self-worth. The twelve items selected from the original for use in the NLSY79 assessment translate into two subscores, a global self-worth score and a scholastic competence score. There is no overall self-perception score. The global self-worth score is a summation of the six "even-numbered" items, beginning with the second item. The scholastic competence score is a summation of the odd numbered items, beginning with item one. These two scales represent two of six subscales developed by Susan Harter. A full description of all the subscales may be found in the SPPC Manual (Harter, 1985). The NLSY79 testing protocol for this assessment may be found in the *Child Supplement*.

The assessment, titled "What I Am Like" in the *Child Supplement*, was, through 1994, completed by children age eight and over. Beginning with 1996, administration is limited to children 12 and over. Scale items are typically phrased as follows:

"Some kids like the kind of person they are BUT Other kids often wish they were someone else."

Children select which option is most like them and then indicate whether that statement is *sort of true* or *really true* for them. Users should note that, with Harter's consultation, very minor wording changes were made to the original items when adapting them for use in the NLSY79 Child. For example, two items each from the Scholastic Competence and Global Self-Worth subscales show the following variation:

Harter wording	NLSY79 Child wording
Some kids often forget what they learn BUT Other kids **can** remember things easily.	Other kids remember things easily
Some kids often forget what they learn BUT Other kids don't do well at their classwork.	Other kids don't do **very** well at their classwork

Some kids are **usually** happy with themselves as a person BUT Other kids are often not happy with themselves.	Some kids are happy with themselves as a person
Some kids are not happy with the way they do a lot of things BUT Other kids think the way they do things is fine.	Some kids are not **very** happy the way they do a lot of things

A value of 4 for each item denotes the highest level of self-worth and a 1 denotes the lowest level.

In the NLSY79, interviewers directly administer this instrument to the children. The interviewer reads each statement to the children, then asks "which kind of kid" they were more like, and follows up by asking whether or not the particular response is "really true for you" or "only sort of true for you." Older children have the option of reading along on printed cards and simply answering whether they are more like the "X" side or the "Y" side of the card.

Scoring the SPPC. Each of the two subscales include six items that are scored between one and four, with higher scores representing greater scholastic competence or greater global self-worth. Only raw scores, which are a simple summation of the six individual items in each scale, are provided, as no national norms are available. Subscore identification may be found in Table 9 and Table 10.

For a small number of cases, there are some missing items. In these instances, a prorated score is computed, assigning average values to the missing items. Flag variables are included in each year's data that identify the degree to which cases have been prorated. For example, a value of zero on these flags indicates that all items were completed and no proration performed; a "1" indicates that one item was missing, and so on.

Completion Rates. As may be noted in Table 12, this assessment has a relatively high completion rate in the current round of about 87.5 percent, with only modest ethnic or racial variability. However, there is evidence that younger children, those under ten (who had been administered this assessment in the pre-1996 survey years), may have had greater difficulty in understanding some of the items. For this reason, scores for younger children may have been

somewhat less reliable and valid. In this regard, it is useful to note that within and cross-year correlations between the two SPPC subscales and the various other cognitive assessments are significantly higher for children age ten and over then for eight and nine year olds. The zero-order correlation between the two subscales is about .3 for eight and nine-year-olds compared with .4 for children age ten and over (Baker, et al., 1993: 130-131). For younger children, there is little association between the two scores and demographic or socioeconomic priors (Mott, et al., 1995).

Validity & Reliability. In general, the reported reliabilities for the NLSY79 administration of these two subscales are somewhat lower than those reported by Harter (1985, 1990). She reports internal reliability of about .8 compared with .67 for the NLSY79 samples. This may partly reflect differences between the samples in their racial, ethnic, or socio-economic mix. Both the *NLSY79 Child Handbook* and *The NLSY Children 1992* include more extensive evaluations of the reliability and validity for these two subscores and the *NLSY79 Child Handbook* includes a review of other literature on the topic.

As a final note, it appears that there has been some escalation in the scores of the Global Self-Worth assessment over time. For example, in 1988, 58.4 percent of the children scored 20 or over, compared with about 63-64 percent in 1990-1992, 69 percent in 1994 and 76 percent (for children age12 and over) in 1996 (see Table 5.3 in the *The1996 NLSY79 Child Assessments: Selected Tables*"). This category exceeded 71 percent in 1998 (see Table 5.4 in the *1998 Child Assessment Tables*) The proportion with very low scores declined during the period. The reason for this remains unclear.

Wechsler Intelligence Scale for Children - Memory for Digit Span

The Memory for Digit Span assessment, a component of the Wechsler Intelligence Scales for Children – Revised (WISC-R), is a measure of short-term memory for children aged seven and over (Wechsler, 1974). The WISC-R is one of the best normed and most highly respected measures of child intelligence (although it should be noted that the Digit Span component is one of the two parts of the Wechsler scale not used in establishing IQ tables).

There are two parts to the Memory for Digit Span assessment. First, the child listens to and repeats a sequence of numbers said by the interviewer. In the second part, the child listens to a sequence of numbers and repeats them in reverse order. In both parts, the length of each sequence of numbers increases as the child responds correctly. In 1996 and 1998, this assessment has been administered to all children age seven through 11 years. In prior rounds, it was typically administered to all children seven and over who had not previously received the assessment, and all ten and eleven year olds (see Table 11).

The child is instructed to repeat a series of 14 numbers (with increasing numbers of digits) forward and a different series of digits in reverse order. Each correct response is worth one point; the theoretical maximum on each of the subscores is, thus, 14 and for the total score, 28. The forward sequence is completed prior to the backward digit sequence. However, entry into the reverse sequence is not contingent on successful entry or completion of the forward sequence. Where appropriate, this assessment is administered in Spanish.

The Digit Span Scores. This assessment generates three non-normed "raw" scores and one overall age-appropriate normed (standard) score. Whereas the normed scores for the other assessments are based on a mean of 100 and a standard deviation of 15, the Digit Span assessment is normed against a distribution which has a mean value of 10 and a standard deviation of 3. Norms are only available for the total score. The variables to be accessed for

these raw and normal scores may be found in Table 9 and Table 10. The norms may be found in the *WISC* manual (Wechsler, 1974: 118-150). The precise instruction and items used in this assessment may be found in the Memory for Digit Span section of the 1996 *Child Supplement*.

Completion Rates. The completion rate for Digit Span is about 90 percent (Table 12) with only limited racial or ethnic variability. Based on a cross-year examination of Digit Span scores, it is difficult to generalize about racial or ethnic differences in scores. In 1998, white scores appear to be slightly above those of the minority groups, with this difference being most prevalent on the "Backwards" assessments (Tables 6.1 through 6.4 in the 1998 Table Report). However, in at least several prior years, different patterns were in evidence, with racial differences not following any generalizable systematic pattern.

As noted above, a Spanish version is available for administration. While this version is available for use, it may be that some Hispanic children with a less than adequate understanding of verbal English (the assessment is verbally administered by an interviewer) nonetheless completed the English version with less than optimal results

Validity & Reliability. In multivariate analyses carried out with the 1992 data which controlled for a wide range of demographic and socio-economic antecedents, the scores of neither black nor Hispanic children were below those of non-Hispanic children white youth on either the forward or backward assessment (the NLSY79 children 1992). In a somewhat different vein, in the same analyses, it was also found that the Digit Span subscores in 1986, in particular the backward assessment, are very useful independent predictors of all of the PIAT scores for older children in 1992. Users wishing more detailed information about the reliability and validity of these assessments as well as a brief discussion of other literature about studies that have used these assessments should consult the *NLSY Child Handbook* and *The NLSY Children 1992* (Baker, et al., 1993; Mott, et al., 1995).

PIAT Mathematics

The Peabody Individual Achievement Test (PIAT) is a wide-range measure of academic achievement for children aged five and over which is widely known and used in research. It is among the most widely used brief assessment of academic achievement having demonstrably high test-retest reliability and concurrent validity. The NLSY79 *Child Supplement* includes three subtests from the full PIAT battery: the Mathematics, Reading Recognition, and Reading Comprehension assessments. Many of the comments related here to the PIAT math subtest are equally appropriate for the other PIAT (as well as PPVT) assessments.

The PIAT Mathematics assessment protocol may be found in the 1998 *Child Supplement*. This subscale measures a child's attainment in mathematics as taught in mainstream education. It consists of eighty-four multiple-choice items of increasing difficulty. It begins with such early skills as recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. Essentially, the child looks at each problem and then chooses an answer by pointing to or naming one of four options.

Basal & Ceiling. The PIAT Mathematics assessment was administered to all children below young adult age whose "PPVT age" was five years and above. Administration of this assessment is relatively straightforward. Children entering the assessment at an age-appropriate item (although this is not essential to the scoring) and establish a "basal" by attaining five consecutive correct responses. If no basal is achieved then a basal of 1 is assigned (see PPVT). A "ceiling" is reached when five of seven items are incorrectly answered. The non-normalized raw score is equivalent to the ceiling item less the number of incorrect responses between the basal and the ceiling.

PIAT Norms. For a precise statement of the norm derivations, the user should consult the *PIAT Manual* (Dunn and Markwardt, 1970, pp. 81-91, 95). In interpreting the normed

scores, the researcher should note that the *PIAT assessments used in the NLSY79 Child were*

normed about 30 years ago. Social changes affecting the mathematics and reading knowledge of

small children in recent years have undoubtedly altered the mean and dispersion of the reading

distribution over this time period. In this regard, a revised version of the PIAT ("PIAT-R") was

released in 1986, too late to incorporate as a 1986 child assessment. To date, we have opted to

maintain internal continuity within the NLSY79 by continuing to use the 1968 version of the

PIAT.

In 1998, the overall (weighted) standard score mean for NLSY79 children completing the

PIAT Mathematics assessment is about 104 compared with 100 for the 1968 norming sample

(1998 tabulations for the PIAT Mathematics test may be found in Tables 7.1 through 7.5 in the

"*1998 Selected Tables*"). Thus, even though NLSY79 children are somewhat disadvantaged

compared with a full cross-section of contemporary American children, they nonetheless score

above average compared to what one might expect from a full national cross-section. It is likely

that this pattern at least partly reflects changes that have occurred in American society in the last

30 years. For example, it is very possible that factors such as child educational television

viewing patterns or involvement in pre-school programs have improved younger children's

readiness for mathematics and reading, if not their advanced capability.

Normalized percentile and standard scores are derived, on an age-specific basis, from the

child's raw score. The user is reminded that *a child's age determination for this assessment is*

based on a PPVT age. The norming procedures are essentially a two-step process with the

percentile scores being derived from the raw scores and the standard scores from the percentile

scores. The reference numbers for the 1986 through 1998 raw and normed scores are listed in

Table 9 and Table 10. The norming sample has a mean of 100 and a standard deviation of 15.

Completion Rates. The majority of the invalidly skipped items in the PIATs between 1986 and 1992 (years when the survey was administered by paper and pencil) fall into two categories. First, there are some children who were inadvertently skipped over even though they were of an appropriate age. Second, a number of children could not be scored because the scoring decision rules were not followed properly so either a basal or ceiling could not be obtained. This looser data collection procedure, which resulted in children being asked a greater number of questions then was required by protocol is no longer utilized beginning with the 1994 assessment administration. The utilization of computer-assisted personal interview (CAPI) techniques introduced with the 1994 child data collection round takes the decision making regarding basal and ceiling procedures out of the hands of the interviewer. Thus, a PIAT assessment can no longer be inadvertently terminated because an interviewer errs in deciding whether a basal or ceiling has been reached.

Users of the PIAT assessments are encouraged to carefully examine the individual response patterns as well as the reasons for invalid scores, particularly for the 1986 – 1992 period. Having the individual responses will permit the user to note that a number of assessments originally considered "unscorable" were scorable once the actual patterns of response on the various assessments were individually considered. This edit was possible because the interviewer recorded the actual response as well as a score of correct or wrong for each answer. Thus, if the correct-wrong item was inadvertently left blank (something which was possible only with paper-and-pencil administration), but the actual response was available, it was frequently possible, in scoring the 1986 through 1992 assessments, to make a post hoc determination of "correctness." In addition, depending on the user's research intention, it may be possible to "score" additional cases if one is willing to sacrifice some precision in the scoring. For example, some additional cases could be scored, if one is willing to accept as adequate a

score that does not deviate by more than one or two point from the "true" score.   This issue became less relevant beginning with the CAPI interviewing procedures in 1994.

As may be noted in Table 12, the overall completion rate for PIAT Math in 1998 is about 89 percent.   There are only modest differences between the white, black and Hispanic completion rates.   This is a lower completion rate that was in evidence for any prior survey round, and largely reflects the lower interior completion status of their mothers.

Changes in Scoring.   Beginning with 1990, several improvements have been introduced into the PIAT norming scheme that should improve the utility of these measures as well as simplifying their use.   First, children between the ages of 60 and 62 months (for whom no normed percentile scores had been previously available) are now normed using percentile scores designed for children enrolled in the first third of the kindergarten year—the closest approximation available to ages 60 to 62 months.

Starting in 1994, children with raw scores translating to percentiles below the established minimum were assigned percentile scores of one; children with raw scores translating to percentile scores above the maximum are assigned percentile scores of 99.   In prior years, the "out-of-range" children had arbitrarily been assigned scores of 0, which led to some inadvertent misuse of the data. (Prior to the 1994 period, children more than 217 months of age are assigned normed scores of -4, since they are beyond the maximum ages for which nationals normed scores are available.)

Validity & Reliability.   In general, the PIAT Math is a highly reliable and valid assessment.   As described in the *NLSY Child Handbook* and *The NLSY Children 1992*, it is closely correlated with a variety of other cognitive measures.   It is both predicted by and predicts scores on a variety of the other assessments.   A particularly strong analytical advantage derived from all of the PIAT assessments is the fact that they have now been repeatedly asked of children

age five and over. Indeed, there are some children age 13 or 14 years in 1994 through 1998 who

have completed these assessments five times. Additionally, most of the children in the Young

Adult sample have in their NLSY79 history several PIAT administrations. This permits one to

carefully examine developmental profiles in relation to school and early-career development.

PIAT Reading Recognition

The Peabody individual Achievement Test (PIAT) Reading Recognition subtest, one of

five in the PIAT series, measures word recognition and pronunciation ability—essential

components of reading achievement. Children read a word silently, then say it aloud. PIAT

Reading Recognition contains eighty-four items, each with four options, which increase in

difficulty from preschool to high school levels. Skills assessed include matching letters, naming

names, and reading single words aloud.

To quote directly from the PIAT manual, the rationale for the reading recognition subtest

is as follows:

> "In a technical sense, after the first 18 readiness-type items, the general
> objective of the reading recognition subtest is to measure skills in translating
> sequences of printed alphabetic symbols which form words, into speech
> sounds that can be understood by others as words. This subtest might also
> be viewed as an oral reading test. While it is recognized that reading aloud
> is only one aspect of general reading ability, it is a skill useful throughout
> life in a wide range of everyday situations in or out of school" (Dunn and
> Markwardt, 1970, pp. 19-20). The authors also recognize that "performance
> on the reading recognition subtest becomes increasingly confounded with
> the acculturation factors as one moves beyond the early grades."

This assessment is administered to children below young adult age whose PPVT age is

five and over. The scoring decisions and procedures are identical to those described for the

PIAT Mathematics assessment, and a description of the process and recognition words may be

found in the 1998 *Child Supplement*. The only difference in the implementation procedures

between the PIAT Mathematics and PIAT Reading Recognition assessment is that the entry

point into the Reading Recognition assessment is based on the child's score in the Mathematics assessment, although entering at the correct point is not essential to the scoring.

As with the PIAT Mathematics, it is important to note that the norming sample was selected and the norming carried out in the late 1960s. This has implications for interpreting the standardized scores of the children in the NLSY79 sample, as already described in the PIAT Mathematics discussion. In this regard, the 1998 NLSY79 sample which has completed the Reading Recognition assessment have a mean standard score of about 107 compared with 100 for the national norming sample (see Table 8.5 in CHRR, 1999).

As with the Mathematics assessment, children with invalid scores on this assessment for the most part never entered the assessment or else were unscorable because of premature termination was more common. In some instances, a careful examination of the individual responses in conjunction with an examination of the interviewer's actual scoring calculations permitted clarification, and ultimately scoring, of additional cases.

It is however, important to note that whereas the actual answer to each item was coded for the PIAT Mathematics responses, this was not done for the PIAT Reading Recognition items. This is one reason why, historically, the overall response rate is slightly lower on the PIAT Reading Recognition assessment: in contrast with the PIAT Mathematics assessment, it was not possible to rectify inadvertent skips for some children on the PIAT Reading Recognition assessment where the "correct-non-correct" check item was inadvertently left blank. Researchers who plan to use the PIAT Reading Recognition assessment extensively are encouraged to examine the individual response patterns. Where a particular researcher does not require great precision on this particular outcome (e.g., a categorization of scores into a number of discrete categories being sufficient), it may be possible to reduce the non-completion rate. In a number of cases, while an exact score may not be determined, an appropriate score

determination (e.g., within two or three points, or a score of at least a certain level) may be possible.

Scoring Changes. The user should note that several improvements were introduced beginning with the 1990 PIAT norming scheme that improved the utility of these measures as well as simplified their use. First, children between the ages of 60 and 62 months (for whom no normed percentile scores had been available in 1986 or 1988) are now normed using percentile scores designed for children enrolled in the first third of the kindergarten year—the closest approximation available to ages 60 to 62 months.

Starting in 1994, children with raw scores translating to percentiles below the established minimum are now assigned percentile scores of one; children with raw scores translating to percentile scores above the maximum are assigned percentile scores of 99. In prior years, the "out-of-range" children had arbitrarily been assigned scores of 0, which led to some inadvertent misuse of the data. (Through 1994, children more than 217 months of age are assigned normed scores of -4 since they are beyond the maximum ages for which national normed scores are available.)

PIAT Reading Recognition Scores. Three scores are reported for the PIAT Reading Recognition assessment for 1986 through 1998: an overall non-normed raw score and two normed scores--a percentile score and a standard score. The norming sample has a mean of 100 and a standard deviation of 15; these were normed against standards based on a national sample of children in the United States in 1968. The reference numbers of the 1986 through 1996 scores may be found in Tables 8 and 9.

Completion, Validity & Reliability. T he overall PIAT Reading Recognition completion rate is, as was true for PIAT Mathematics, the lowest completion rate yet reported for this assessment, and largely reflects continuing declines in interview completion rates for their

mothers. About 89 percent of eligible children completed the assessment, with little difference between non-Hispanic white and minority children (Table 12). As is true for the mathematics assessment, the recognition assessment is considered quite reliable and valid. The *NLSY Child Handbook* includes a comprehensive discussion of these issues, drawing on material from the PIAT Manual as well as a variety of research which has been completed using the NLSY79 Child PIAT reading data. This discussion also includes internal CHRR evaluation of the cross-year correlations with other NLSY79 PIAT scores as well as the full spectrum of other cognitive assessments. Additionally, *The NLSY Children 1992* synthesizes evidence of strong longitudinal independent associations between PIAT reading and a full set of demographic and socio-economic priors. In general, this assessment, as well as all of the other Peabody assessments, is widely used and has a well-established academic track record.

PIAT Reading Comprehension

The Peabody Individual Achievement Test (PIAT) Reading Comprehension subtest measures a child's ability to derive meaning from sentences that are read silently. For each of 66 items of increasing difficulty, the child silently reads a sentence once and then selects one of four pictures that best portrays the meaning of the sentence.

> "While understanding the meaning of individual words is important, comprehending passages is more representative of practical reading ability since the context factor is built in, which plays an important role, not only in deciphering the intended meaning of specific words, but of the total passage. Therefore, the format selected for the reading subtest is one of a series of sentences of increasing difficulty. The 66 items in Reading Comprehension are number 19 through 84, with item 19 corresponding in difficulty with item 19 in Reading Recognition." (Dunn and Markwardt, 1970, pp. 21-22).

The PIAT Reading Comprehension assessment is administered to all children below young adult age whose PPVT age is five years and over who scored at least 19 on the Reading Recognition assessment. (NOTE: From 1986 through 1992, PIAT Reading Comprehension was

actually administered to all children who scored 15 or higher on Reading Recognition. This lowered threshold was used to maximize our ability to ultimately score the Reading Comprehension assessment for those cases where interviewers made minor addition errors in totaling the Reading Recognition test, computing actual scores of 19 or more as only being 15 through 18.)

Children who score less than 19 on Reading Recognition are assigned their Reading Recognition score as their Reading Comprehension score. If they score at least 19 on the Reading Recognition assessment, their Reading Recognition score determines the entry point to Reading Comprehension. Entering at the correct location is, however, not essential to the scoring. Basals and ceilings on PIAT Reading Comprehension as well as an overall non-normed raw score are determined in a manner identical to the other PIAT procedures. The only difference is that *children for whom a basal could not be computed (but who otherwise completed the comprehension assessment) are automatically assigned a basal of 19*. Administration procedures are described in the 1996 *Child Supplement*. As with the other PIAT tests, norming was accomplished in the late 1960s with all its attendant potential analytical problems. These are noted in more detail in the PIAT Mathematics discussion. For a precise statement of the scoring decisions and the norm derivations, the user should consult Dunn and Dunn (1981) and Dunn and Markwardt (1970).

The PIAT Reading Comprehension Scores. The NLSY79 Child dataset provides the following PIAT Reading Comprehension scores: overall non-normed raw scores that can range from zero to 84, normed percentile scores, and normed standard scores. Reference numbers for the 1986 through 1998 scores may be found in Tables 8 and 9. It should be noted that many younger children (age seven years and below) who receive low raw scores *cannot be given normed scores* because their scores are out of range of the national PIAT sample used in the

norming procedure. These children have been assigned -4 codes on the percentile and standard score variables. Researchers wishing to keep these children in their analyses will thus need to consider special decision rules. The way to identify these children, of course, is to cross-classify children by their raw score and standard score. They will be identified by having a raw score of zero or greater but a standard and percentile score of -4.

If one is using the PIAT Reading Comprehension assessment for analyzing five- and six-year-olds, the proportion of children without a standard score is a major constraint that cannot be ignored. A large proportion of five- and six-years-olds that have a *valid* raw score on Reading Comprehension could not be given a normed score. All of these children had raw scores below 19 and thus had their Reading Recognition score imputed as the Comprehension score; one solution for the youngest children (those with PPVT ages under 7) is to limit analyses to Reading Recognition.

By applying procedures parallel to those used with PIAT Mathematics, it was sometimes possible to clarify the score of a previously "unscorable" child by carefully examining the individual response patterns, particularly where the actual response for the "correct-incorrect" item had not been completed. This was more relevant in the 1986 – 1992 "pre-CAPI" administration survey rounds. In this way, we were able to retrieve a number of cases not previously scorable. Depending on a researcher's individual inclination or need for precision, it may be possible to score, in an approximate manner, a number of additional children. In order to accomplish this, the researcher will need to examine the individual PIAT comprehension items. Researchers who plan to use this outcome extensively are encouraged to examine the individual item responses.

Scoring Changes.   Several changes were introduced beginning with the 1990 PIAT norming scheme in order to improve the utility of these measures and to simplify their use. First,

children between the ages of 60 and 62 months (for whom no normed percentile scores had been previously available) are now normed using percentile scores designed for children enrolled in the first third of the kindergarten year—the closest approximation available to ages 60 to 62 months.

As of the 1994 round, children with raw scores translating to percentiles below the established minimum are now assigned percentile scores of one; children with raw scores translating to percentile scores above the maximum are assigned percentile scores of 99. In prior years, the "out-of-range" children actually had arbitrarily been assigned scores of 0, which led to some inadvertent misuse of the data. (Prior to 1994, children more than 217 months of age are assigned normed scores of -4 since they are beyond the maximum ages for which normed scores are available.)

Completion Rates. Reading Comprehension completion rates have typically been quite low. For example, in 1992 only about 86 percent of eligible youth received a comprehension score. In the pre-1994 survey period, several reasons have been suggested for the relatively low comprehension completion rate. In some instances, the assessment was simply skipped over with no reason given. In other instances, a valid Reading Recognition score was available but the interviewer neglected to assess the child on Reading Comprehension. More typically, the Reading Comprehension assessment was attempted, but the interviewer did not attempt a sufficient number of items to attain a basal or ceiling. An apparently common problem was where an interviewer entered Reading Comprehension at a fairly low level, apparently tested a child, but did not record all of the responses. As with all of the assessments, the researcher is encouraged to examine the scoring patterns for the invalid responses. Depending on one's research objectives, some flexibility in re-scoring may be possible.

The PIAT Comprehension completion rates in 1994 and 1996 are substantially higher then in 1992, almost reaching 90 percent, but have dipped to about 88 percent in 1998, reflecting the lower overall survey completion rates. This decline appears for virtually all ages, with the poorest completion rates continuing to be in evidence for the youngest, 5-6 year old children.

Validity & Reliability. As with the other PIAT assessments, Reading Comprehension is generally considered to be a highly reliable and valid assessment that has been extensively used for research purposes. This version was normed in the late 1960s and thus is subject to the same analytical constraints as the other PIAT assessments. In this regard, while the level of the standardized scores appears too high, it is likely that the patterning of the responses is reasonable. That is, higher scores will represent better outcomes in comparison with lower scores. Readers wishing additional detail regarding specific research, which has utilized this NLSY79 assessment, should examine the PIAT discussion in the *NLSY Child Handbook* and obtain the most recent NLSY79 Child Assessment Research Bibliography from the CHRR. Additional information documenting the close association between PIAT Comprehension and a full range of socio-economic and demographic maternal and family antecedents may be found in *The NLSY Children 1992* (Mott et al., 1995). Finally, 1998 scores are summarized in Tables 9.1 through 9.5 in the "*1998 Selected Tables*."

The Peabody Picture Vocabulary Test - Revised (PPVT-R)

The final child assessment in the NLSY79 is the Peabody Picture Vocabulary Test (PPVT). "The PPVT-R measures an individual's receptive (hearing) vocabulary for Standard American English and provides, at the same time, a quick estimate of verbal ability or scholastic aptitude" (Dunn and Dunn, 1981). This assessment can be given to all children age three and over. The PPVT-R assessment protocol may be found in the *Child Supplement*. For the actual

images (or "plates") presented to the child, one must access the PPVT-R Manual and materials (Dunn and Dunn, 1981). The English language version of the assessment consists of 175 vocabulary items of generally increasing difficulty. The child non-verbally selects one of four pictures which best describes a particular word's meaning. A child's entry point into the assessment is based on his or her PPVT-R age.

Scoring Procedures. Children enter the assessment at an age-appropriate level, although this is not essential to the scoring. A "basal" is established when a child correctly identifies eight consecutive items. (Exceptions to this are those cases where a basal cannot be established. In these instances, a child is given a basal of one.) A "ceiling" is established when a child incorrectly identifies six of eight consecutive items. A child's raw score is determined by adding the number of correct responses between the basal and ceiling to the basal score.

Norms. The PPVT-R was standardized on a nationally representative sample of children and youth. The norming sample included 4200 children in 1979, and norms development took place in 1980 (Dunn and Dunn, 1981). For a comprehensive discussion of this norming procedure, researchers should refer to the *PPVT-R Manual for Forms L and M* (Dunn and Dunn, 1981). The *PPVT-R Manual* provided information about the linkage between the standard and percentile score.

Users may note one very important distinction between the PPVT-R and PIAT scores—a difference of particular interest to those who plan to concurrently use both assessments. Whereas the PIAT assessments had surprisingly high mean scores (see PIAT discussions) for a sample which includes an above average proportion of disadvantaged children, the PPVT-R means are somewhat below those of the norming sample. The NLSY79 PPVT-R sample has a mean standard score of about 97 and standard deviation of about 20. Only the white sample had a mean approximating the overall national average of 101 (see Table 10.4 in the "*1998 Selected*

Tables"). This large differential between the NLSY79 PIAT and PPVT-R mean scores at least

partly reflects the fact that the PPVT-R norming sample is relatively contemporary (1979),

whereas the PIAT norming sample is from the late 1960s. The reader is referred to the *NLSY79*

Child Handbook for a more comprehensive evaluation of racial, ethnic and, socio-economic

differentials in PPVT-R scores using the 1992 NLSY79 data which included PPVT-R assessment

scores for all children 3 and over.

In 1986, this assessment was only given in English. However, beginning in 1988, a small

number of children who preferred to do so were given the Spanish version of this assessment, the

"Test de Vocabulario en Imágenes Peabody." For this reason, post-1986 assessment results may

be less culturally biased than the 1986 version.

In 1986, all children age three and over were given this assessment. In 1988, all ten- and

eleven-year-olds (our "index" population) as well as other children age three and over who had

not previously completed the assessment in 1986 were given this assessment.   In 1990, all

children age ten and eleven as well as all other children age four and over who had not

previously completed the assessment were eligible for the PPVT-R assessment.   In the 1992

survey round, all children age three and over were eligible to be assessed. Thus, there are at least

two survey points (1986 and 1992) in which all age-eligible children who were still being

interviewed had a PPVT-R score. Of course, many of these children may also have had an

intervening (at age 10 or 11) PPVT-R score. In 1998, as in several preceding rounds, the PPVT

administration was largely limited to 4 and 5 year-old children who had not been previously

administered the test as well as the 10-11 year old "index" population. As has been reported for

several preceding data collection waves, the youngest children administered this test score the

poorest, probably reflecting their unfamiliarity with a testing environment. Their lower scores do

not reflect lower status as these younger children have parents with more education then do the

older, 10-11 year olds. In general, as with the PIAT assessments, overall completion rates, as well as age-specific rates are down significantly from 1996. The across-year administration pattern is described in Table 11.

As with PIAT Math and Reading Comprehension, it was possible, primarily in the pre-CAPI years, to improve the overall quality and completion level by utilizing information on the actual responses where "correct-wrong" check item had inadvertently been skipped. In addition, depending on the user's research intention, it may be possible to "score" additional cases if one is willing to sacrifice some precision in the scoring. For example, some additional cases could be scored if one is willing to accept as adequate a score that does not deviate by more than one or two points from the "true" score. For a precise statement of the scoring decisions and some of the norm derivations, the user should consult the PPVT-R Manual (Dunn and Dunn, 1981, pp. 96-110, 126).

Norms. Beginning in 1990, the procedure used to create the NLSY79 Child PPVT-R normed scores was refined in two important ways. First, children with raw scores that translated into standard scores between 20 and 39 are now being normed using the *PPVT-R Supplementary Norms Tables* (American Guidance Service, 1981). Second, raw scores that would translate to normed standard scores above the maximum provided are now assigned standard scores of 160, and raw scores translating to standard scores below the minimum are now assigned standard scores of 20. In prior years, these children had been assigned a standard score of zero. Three 1986 through 1996 scores are provided for this assessment for each child; a non-normed raw score, a normed standard score, and a normed percentile score. The reference numbers for these items can be found in Tables 8 and 9. Instructions in the *PPVT-R Manual* provide information about the linkage between the raw score and the standard score, and the percentile score is mechanically determined by the known linkage between the standard and percentile. The

NLSY79 Child sample has been normed against a national population with a standard score mean of 100 and a standard deviation of 15.

The user is reminded that the eligibility of children for the PIAT and PPVT-R assessments is based on their "PPVT-R age," which can differ from their calendar age (in months). This was elaborated on earlier in this Handbook. When working with the PPVT-R or PIAT assessments, the "PPVT-R age" variable should be used.

Validity & Reliability. The PPVT is among the best-established indicators of verbal intelligence and scholastic aptitude across childhood. It is among the most frequently cited tests in Mitchell's (1983) "Tests in Print." Numerous studies have replicated the reliability estimates from the PPVT's standardization sample. The *NLSY Child Handbook* synthesizes much of this work. This report also provides cross-year (1986-1990) reliability and validity evaluation using the NLSY79 Child data. *The NLSY Children 1992* contains an evaluation of the quality issues for the 1992 PPVT sample, which included the full spectrum of children age three and over. These analyses show strong associations between a full range of social and demographic priors and 1992 PPVT scores. The report also documents strong independent linkages between PPVT scores in 1986 and PPVT, PIAT Reading and Mathematics, and SPPC scores in 1992. Typically, stronger associations are found for white and Hispanic then for black children. One other finding of importance should be mentioned. More than for any of the other assessments, substantial racial and ethnic variations may be noted for the PPVT. The average non-Hispanic white child scores at the 52nd percentile compared to the 30th percentile for his or her Hispanic counterpart and the 25th percentile for his or her black counterpart (see Table 10.3 in CHRR, 1999b *The 1998 NLSY79 Child Assessment: Selected Tables*). Substantial ethnic and racial variations remain in multivariate analyses even with demographic and socio-economic controls.