

6. TWO BAGS TASK IN THE ECLS-B

This section discusses the rationale for the transition from the Nursing Child Assessment Teaching Scale (NCATS) at 9 months to the Two Bags Task at 2 years. This is followed by a description of the in-home administration of the Two Bags Task, as well as the quality control procedures that were undertaken to ensure that the data obtained were of the highest quality, including training the interviewers, training the trainers on the coding system, and training the coders. In addition, a summary of coder reliability and Cronbach's alpha for the Two Bags subscales is presented, followed by a comparison of the 2-year Two Bags Task rating scales with the 9-month NCATS scale scores and descriptives for the Two Bags Task by key demographic grouping variables.

6.1 Technical Review Panel Advice

The design of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) was guided by three principles. The first guiding principle was to obtain measures of growth through repeated measures at multiple time points. The second was to obtain, wherever possible, direct measures of child functioning rather than to rely on parental reporting in order to reduce potential response bias. The third guiding principle was to obtain information about a broad spectrum of children's early experiences in order to understand their relationship to children's development over time.

Consideration of the above principles led to the decision, strongly endorsed by Technical Review Panel (TRP) advisors, to include a direct measure of parent-child interaction. Parent-child interaction is a key aspect of children's early experiences known to predict subsequent child outcomes. To capture the full breadth of young children's functioning, it was important to include a direct measure of their socioemotional functioning. During infancy and toddlerhood, socioemotional functioning is easiest to assess during mother-child interaction, because this interaction provides a context within which the child's emotional functioning can be elicited. This approach is supported by multiple lines of research in developmental psychology, including attunement (Stern 1985), intersubjectivity (Trevarthen and Aitken 2001), social referencing (Walker-Andrews 1998), and emotion regulation (Miller et al. 2002).

TRP advisors advocated the inclusion of a direct observational measure of the parent's and child's behaviors in order to obtain information about the quality of children's interactions that may

influence their readiness for school. TRP members further advised that the NCATS of the Nursing Child Assessment Satellite Training (NCAST) would be useful for the early rounds of the ECLS-B for several reasons. It was used in the Early Head Start Research and Evaluation project, a study with a large sample of very young children, enabling a comparison of the results from the two studies; it has a standardized training that would ensure consistency of coding; it could be used at both the 9- and 18-month data collections, which would further the goal of obtaining repeated measures to examine growth; and it can be coded from videotape, thereby reducing burden to interviewers who would simply videotape the interaction and not code during administration.

A different observational measure of parent-child interaction needed to be implemented at the 30-month and later data collections because the developer of the NCATS did not endorse its use beyond 24 months of age. For the purposes of the ECLS-B 2-year data collection, the TRP members suggested the Three Bags Task as a viable measure of parent-child interaction, because it is one of the few coding systems that can be used in large-scale studies, has excellent training materials, good psychometric properties, and, while brief, produces robust scores predictive of later growth in both cognitive and socioemotional domains. As design of the 30-month data collection progressed, it was decided to implement the Three Bags Task, which has been used with success in other large-scale studies, including the Early Head Start Research and Evaluation Project (with a national sample of approximately 3,000 very young children) sponsored by the Administration on Children, Youth, and Families (ACYF), and the National Institute of Child Health and Human Development (NICHD) Early Child Care Study, a project involving a consortium of academic and social policy researchers (NICHD Early Child Care Research Network 2004, 2005).

The Three Bags Task is a semi-structured activity completed by the parent and child in interaction. Because the Three Bags Task requires at least 15 minutes for the dyad to complete, the activity was shortened to just two bags/activities, which could be completed in 10 minutes, with the provision that one activity should be a joint book reading activity. During this 10-minute task, the parent-child dyad is asked to play with two different sets of toys, each placed within a separate numbered bag. In the 2-year ECLS-B, bag number 1 contained a set of dishes, and bag number 2 contained a children's picture book, *Good Night, Gorilla*, by P. Rathmann (1994). The dyad was told that they had 10 minutes to play with the two bags, the only restriction being that they had to play with the bags in numerical order. The parent and child were videotaped while they engaged in the activities. The videotapes were sent back to Westat, where staff trained on the rating scale rated the parent on six global scales and the child on three global scales.

The Two Bags Task rating scales include six parent rating scales and three child rating scales. The scales are on a 7-point Likert-type rating scale that ranged from very low (1) to very high (7). Each rating level is well described in the coding manual with specific examples to illustrate the concept and target behaviors.

To code a videotape, the coder watched the videotape and observed the target behaviors, making notes that would help rate the items. When the videotape was finished, the coder rated each item on the basis of observations made while watching the videotape.

The six parent rating scales include the following:

- Parental Sensitivity (variable name C2SENSTV): This scale focuses on how the parent observes and responds to the child's cues (including gestures, expressions, and signals), both when the child is distressed and not distressed. The key defining characteristic of parental sensitivity is that the parent's response is child-centered. Sensitive parenting involves "tuning-in" to the child and manifesting awareness of the child's needs, moods, interests, and capabilities.
- Parental Intrusiveness (variable name C2NTRUSV): This scale reflects the degree to which the parent controls the child rather than recognizes and respects the validity of the child's perspective. Intrusive interactions are adult-centered rather than child-centered and involve imposing the parent's agenda on the child despite the child's protest or defensiveness. Extreme intrusiveness can be seen as over-control to the point where the child's autonomy is minimized or rejected. The key characteristic is that the intrusiveness is seen from the point of view of the child and careful observation of the child's reaction to the intrusiveness is required.
- Parental Stimulation of Cognitive Development (variable name C2COGDEV): This scale focuses on the parent's effortful teaching to enhance perceptual, cognitive, and language development. A stimulating parent is aware of the child's developmental level and aims to bring the child to the next level. If the topic or method of stimulation is not matched to or slightly above the child's developmental level or interest, then the parent's behavior is not seen as stimulating cognitive development.
- Parental Positive Regard (variable name C2POSRGD): This scale assesses the parent's expression of love, respect, and admiration for the child. Positive regard is seen in the way the parent listens, watches attentively, and looks into the child's face when talking to him/her. Parents who give praise without a warm tone as well as those who do not praise when the opportunity presents itself would not receive the highest score.
- Parental Negative Regard (variable name C2NEGRGD): This scale reflects the parent's expression of discontent with, anger toward, disapproval of, or rejection of

the child. The key is to score parental negative regard from the point of view of the child, and it should be scored independently of the parent's positive behaviors captured in the positive regard scale.

- Parental Detachment (variable name C2DETACH): This scale measures the parent's awareness of, attention to, and engagement with the child. This includes both the extent to which the parent interacts with the child (i.e., the amount of interaction) and the way in which the parent interacts with the child (i.e., the quality of interaction). Detachment can take the form of being consistently inattentive, being inconsistently attentive, or interacting with the child in a perfunctory or indifferent manner.

The three scales that assess children's behaviors include the following:

- Child Engagement of Parent (variable name C2ENGPRT): This scale reflects the extent to which the child shows, initiates, and maintains interaction with the parent, and the extent to which the child communicates positive regard or positive affect to the parent. At the higher end of the scale, the child expresses sustained positive affects toward the parent (through smiling, laughter, etc.) and frequently looks at and attempts to interact with the parent. At the lower end of the scale, the child displays no affect with the parent or ignores or overtly rejects the parent.
- Child Sustained Attention (variable name C2STNATT): This scale assesses the child's ability to sustain attention to and involvement with objects. A child low on sustained attention could seem apathetic, bored, distracted, distressed, or aimless while a child high on sustained attention is able to focus attention when playing with an object and appears involved in what he/she is doing.
- Child Negativity Toward Parent (variable name C2NEGPRT): This scale measures the degree to which the child shows anger, hostility, or dislike toward the parent. At the high end, the child is repeatedly and overtly angry with the parent. The important point is that at this age, the child may express negativity toward the parent by hitting an object, the floor, or him/herself by pushing the parent away, by throwing a toy, or by using a negative expression to communicate that he/she wants or does not want something ("No!"). Therefore, the context of the negative expression should be taken into account when determining the extent to which it is directed toward the parent.

6.2 Rationale for Transition to Two Bags Task

The combining of the 18- and 30-month data collections into a single 2-year data collection necessitated a decision about whether to use the NCATS or the Two Bags Task. Had the 18-month data collection gone forward as planned, there would be no question that the NCATS would be included at 18 months and the Two Bags Task at 30 months and at preschool. However, the switch to 2 years shifted the considerations somewhat with regard to obtaining continuity of measurement. For one, a decision had to

be made about which measure, NCATS or Two Bags, would provide the repeated measure for estimating growth over time: NCATS at 9 months and 2 years, and Two Bags at preschool, or NCATS at 9 months only and Two Bags at 2 years and preschool. The decision was made to administer the NCATS at 9 months and the Two Bags at 2 years and preschool.

Several factors contributed to the decision about which combination of assessments to use. First, 2 years is at the upper limit of the age range for which empirical data support the use of the NCATS, according to its developer, Dr. Kathryn Barnard of the University of Washington. She did not encourage using the NCATS beyond that age because most research using the NCATS has been concentrated on children up to about 2 years of age, with relatively fewer studies on children older than 2 years of age. Consequently, it is not clear that the NCATS can reliably measure parent-child interactions for children older than age 2 (i.e., 24 months). The lack of a research base for the NCATS beyond 2 years would not be a problem if all the children in the ECLS-B were seen promptly within the predetermined “ideal window” of 2 years +/- 4 weeks. However, experience during the 9-month collection when some of the children were seen many months later demonstrated that it was unrealistic to expect that all home visits would be completed within this window. Because only one observational measure could reasonably be used at 2 years, the ECLS-B had to select a different measure if there were no empirical support for the NCATS norms beyond 2 years.

Second, the Two Bags task has the advantage that it can use a parent-child joint book reading activity as one of the tasks. This would give the ECLS-B the opportunity to obtain a direct observational measure of mother’s and child’s language use and literacy behaviors, an important consideration for a study examining the aspects of children’s early experiences that prepare them for later school entry and sustained school achievement. Indeed, Hart and Risley (1995), have built a strong argument for the effects of early experiences on children’s later outcomes, in particular, that the amount of time parents spend talking to their children in the early years of life directly influences children’s future school achievement.

The third consideration in choosing an assessment was ease of administration and coding, as well as cost. It is quite expensive to obtain videotaped interactions of parent and child. Several coding systems have been used to code the Two Bags Task; however, they are generally similar and involve global ratings (on a 4- or 7-point scale) of salient aspects of parent and child behavior. Because this coding system is global, each case can be coded in real-time on one pass through the videotape. Total coding time would be about 12–14 minutes for the Two Bags Task, whereas the average coding time for

the NCATS during the 9-month national study was approximately 17 minutes. Coding time per tape was an important consideration for a large-scale study, which could involve coding up to 10,000 videotapes.

In addition, the Two Bags Task is more straightforward and efficient to administer than the NCATS. The parent is handed the two bags and asked to play with the child for 10 minutes. The NCATS, on the other hand, has complicated and sometimes ambiguous instructions in which the parent must review a list of age-appropriate activities and select the first activity that the child cannot do. Task selection is verified by asking the mother if there is another activity after the selected task that the child can do. If there is, then the next task after that is selected. Task selection for the NCATS in the ECLS-B was difficult because often parents selected a task that was too young so that they could be assured that the child would be able to perform it on the videotape, or too old because the parent wanted the child's precocious abilities on videotape. For these administrative reasons, the Two Bags Task would be less burdensome in the field and obtain more reliable information because all children receive the same tasks.

6.3 Two Bags Task Protocol for In-Home Administration

The Two Bags Task is a videotaped interaction. Therefore, interviewers administering the Two Bags Task during the home visit used a handheld video camera to film the parent and child engaging in the two activities that comprise the Two Bags Task. During the national training, interviewers were taught to administer and to videotape the Two Bags Task. The training included extensive practice, emphasizing good filming techniques and skillful use of the camera in conjunction with faithful administration of the Two Bags Task.

The Two Bags Task administration during the home visit was standardized to ensure that all interviewers administered the task in the same way to all parent-child pairs. To ensure this standardization, step-by-step Two Bags Task administration instructions were included in the Child Activity Booklet in a separate tabbed section for the Two Bags Task. These instructions included a verbatim script that was read to the parent. Interviewers also asked parents whether or not they had previously read *Good Night, Gorilla* to their child, and if so, how often. Interviewers were expected to record parents' answers in check boxes on the administration pages in the Child Activity Booklet, making sure to record verbatim answers related to frequency. The interviewer also recorded the start time of the Two Bags Task and the language used by the parent when talking to the child.

In the case of twins, the interviewer administered the Two Bags Task separately for each twin, but recorded both on the same videotape and used the same activities. This introduced the problem of familiarity with the storybook as a confounding variable. It was possible that on the second reading of the storybook, the parent would alter the reading in some systematic way. Therefore, interviewers were instructed to counterbalance the administration of the Two Bags Task to twins. It would not be possible to impose a true counterbalanced design (in which, say, a random number generator was used to determine order of administration to all twin pairs in the ECLS-B before the 2-year data collection began), however, because this would have been too burdensome to field staff and probably not a realistic expectation. Therefore, field staff was instructed to administer the Two Bags Task to the first-born twin on odd-numbered days and to the second-born twin on even-numbered days. Field staff also recorded in the Child Activity Booklet which twin had been administered the Two Bags Task and in what order.

Unlike at 9 months, when a triadic NCATS involving the mother and both twins simultaneously was obtained after the mother completed the NCATS with each twin separately, there was no triadic Two Bags Task. For further information about triadic NCATS activities, please refer to the *ECLS-B Methodology Report for the Nine-Month Data Collection, Volume 1: Psychometric Characteristics* (NCES 2005–100) (Andreassen and Fletcher 2005).

After completion of the home visit, the field representative then sent the Two Bags Task videotape and the Child Activity Booklet, along with other data collection materials, to Westat's home office for receipting and coding by expert coders.

6.4 Two Bags Task Field Staff Training, Trainer Training, and Coder Training

Three different types of training were required for the Two Bags Task. The first was field staff training. Field staff was trained to obtain high-quality videotapes and to administer the Two Bags Task to the parent and child according to standardized procedures. Second was the trainer training. Home office staff targeted to train coders on the Two Bags Task coding system attended a training session at Columbia University Teachers College. The third was coder training held at the home office. Two Bags Task coders participated in extensive training to ensure reliability¹ of coding comparable to Teachers College standards.

¹ Reliability in this case refers to inter-rater reliability, which is the degree to which different raters or observers give consistent ratings to the same observed behaviors from the same videotapes.

Each of these trainings is described in more detail in the following sections. These descriptions are followed by a summary of quality control procedures that were followed to prevent coder drift from the standards as the year of data collection progressed. The final section summarizes how the Two Bags Task performed in the ECLS-B and presents information about inter-lab agreement between the trainers compared with the coding supervisor at Teachers College and intra-lab reliability between the coders and the reliability consensus coding by the Westat coding supervisor and assistant trainers.

6.4.1 Field Staff Training

For the 2-year data collection, some field staff were returnees from the 9-month data collection and some were new to the ECLS-B. Returning field staff already knew how to operate the videocamera used to tape the parent-child interactions. To enable new field staff to become familiar with the videocamera prior to training and thereby reduce the amount of time required during training, an 8 mm videocamera and an 8 mm cassette, together with an accompanying manual, were sent to each new trainee prior to the national training in Los Angeles. The trainees were instructed to follow the instructions in the manual and to practice using the videocamera at their convenience before coming to training. In addition, the field representative manual provided to all trainees included detailed instructions on videotaping and administering the Two Bags Task. Interviewers were able to refer to this manual during the field period as needed.

By the time of the national training, all trainees were familiar with the operation of the videocamera. This enabled attention to be focused directly on the correct administration of the Two Bags Task procedures at the national training. Trainees were instructed to follow the Two Bags Task administration steps verbatim as presented in the Child Activity Booklet. They then administered the Two Bags Task to each other in sets of three, in alternating turns, one playing the role of the interviewer, one the parent, and one the child.

National training did include emphasis on proper videotaping techniques to obtain a high-quality videotape of the Two Bags Task interaction. A high-quality videotape was critical to successful Two Bags Task coding. Therefore, trainees received hands-on practice and extensive feedback about their videotaping. This was done during the sessions involving direct instruction and also during the live-practice session when training staff circulated through the rooms and watched over the shoulders of field

staff as they videotaped their partners administering the Bayley Short Form–Research Edition (BSF-R). Westat staff members reviewing the videotape to score the BSF-R administration also reviewed the quality of the videotaping. Any videotape that was not of sufficient quality (e.g., audio level too low, lighting level too low, camera faced toward a window so that the dyad was seen only in silhouette, etc.) was noted, and the videotaper was required to attend a help session and/or demonstrate good videotaping skills to her or his lead trainer. In this way, all field staff who had trouble producing a high-quality videotape received intervention and retraining before going into the field.

In addition, videotapes from each field staff member were quality reviewed by the Two Bags Task coding staff on an ongoing basis as they were received at Westat. Feedback on videotape quality was given to all field staff within about a week of receipt of the videotapes. For further information about quality control procedures, please see section 6.5.

6.4.2 Trainer Training

The first task of training the trainers was to have them trained on the coding system so that they, in turn, could train the individuals who would actually be coding the videotapes. The trainer training was done by a graduate student working in the laboratory of Dr. Jeanne Brooks-Gunn at Teachers College, together with Christy Brady-Smith, the first author of their coding manual (Brady-Smith et al. 1999). This individual had been the reliability coder for the Three Bags Task used in the Early Head Start Research and Evaluation Study, thereby ensuring that the Westat Two Bags Task trainers would be trained to the same standards as those used in the Early Head Start study and that results would be comparable to that study.

In May 2003, four Westat staff members from the Child and Family Studies area attended the specialized training at Teachers College. The training took place over the span of 3 days and was conducted by the lead coder for the Early Head Start Research and Evaluation Study. Instruction over the first 2 days consisted of a review of the rating scales interspersed with video clips of examples of the types of behaviors in the rating scales. The third day consisted of reliability coding of seven videotapes. To pass the training, the Westat staff members had to code each videotape and score within 90 percent agreement with the reliability coding.

The procedure followed by the Early Head Start Research and Evaluation Project had been that, once an individual had passed the training, inter-lab reliabilities were provided for a maximum of the first 30 videotapes after training, or until the coder could sustain 90 percent agreement with the reliability coding for 5 consecutive tapes, whichever came first. The same procedure was followed for the Westat trainers. After completion of training, each trainer coded reliability tapes provided by Teachers College. Tapes were sent in batches of 5 to Westat. The trainers coded the tapes and sent their coding sheets to CUTC in batches. The actual scoring of the reliability testing for certification was done by the lead Three Bags Task coder at Teachers College. Westat's trainers established reliability on all the Two Bags Task rating scales quickly and only required an average of 12 reliability tapes before becoming reliable at 90 percent agreement for all the rating scales.

One staff member, who had been the coding supervisor for the NCATS during the 9-month data collection, was designated to be the coding supervisor for the Two Bags Task and lead trainer for the training of coders. A second individual co-led this training and was designated to be the back-up for the supervisor. Together, the coding supervisor and this assistant consensus coded² all reliability tapes that were used to establish intra-lab reliability between the coders and the reliability videotapes. A third individual, a member of the Child Development Team, was also called upon to resolve coding questions that arose during the course of the year and also coded incoming videotapes as necessary, depending on the work load. The fourth individual served as the liaison between the Child and Family Studies area and the Two Bags Task coding workshop. The coding supervisor, her assistants, and the liaison were able to establish, maintain, and share with coders a Two Bags Task coding knowledge base that contributed to the maintenance of coding reliability across the data collection year.

6.4.3 Coder Training

All six NCATS coders from the 9-month data collection who were still on staff at the beginning of the 2-year data collection were retained to code Two Bags Task videotapes. These coders had already demonstrated their coding competence with weekly NCATS reliabilities above the 85 percent agreement criterion. An additional coder, recruited from within Westat, completed the observation skills test that had been used in the recruitment of NCATS coders and passed it at greater than 90 percent. The analyst is referred to the *ECLS-B Methodology Report for the Nine-Month Data Collection, Volume 1:*

² Consensus coding in this context, refers to the process where two (or more) individuals each code a videotape independently and then compare their ratings item-by-item. If there are any discrepancies between the ratings, the discrepancies are resolved by discussion. The result is a final set of ratings that can be used as a standard against which to compare the ratings of other coders.

Psychometric Characteristics (NCES 2005–100) (Andreassen and Fletcher 2005) for further information about this observation skills test.

The training for the seven coder trainees was held about 3 weeks after the trainer training. During that interval, the trainers had established their ongoing reliability on the rating scales, prepared the coder training materials tailored to the specific needs of the Two Bags Task in the ECLS-B and obtained videotapes for training and reliability purposes. The coder training took a full 5 days. The first 3 days were devoted to an introduction to the nine Two Bags Task rating scales, interspersed with videoclips of mother-child interactions to demonstrate the target interaction behaviors, with additional coding practice devoted to one rating scale at a time. The fourth day provided practice in coding Two Bags Task videotapes on all the rating scales simultaneously. The culmination was on the fifth day, when trainees completed seven reliability videotapes. All trainees passed these reliability tapes at the required 90 percent agreement or greater. Following training, the supervisor and her assistants provided coding support to the new coders on an as-needed basis. In addition, if a coder encountered a videotape that was difficult to code, it was brought to the attention of the supervisor who conducted a weekly “brown-bag” coding review session to discuss coding issues and difficulties that may have arisen during the week.

Initially, only English videotapes were coded because none of the coders were fluent in any other major language, such as Spanish or Chinese. Therefore, all foreign language videotapes were put aside. After a sufficient number of foreign language videotapes had been assembled, new coders fluent in the required languages were added to the coding staff and two subsequent trainings were conducted on an as-needed basis. The second training for two additional coders, one fluent in Mandarin and one fluent in Spanish, occurred 4 months after the first training, and was led by the coding supervisor and assistant. A third training became necessary when the Spanish-speaking coder resigned, and a new one had to be hired. This training involved only this one Spanish-speaking trainee and the coding supervisor, who followed the same training script and procedures as those used for the other trainings.

6.5 Two Bags Coding Quality Control Procedures and Reliability

In keeping with procedures instituted for the 9-month NCATS coding, coders worked up to 4 hours a day, coding a maximum of 10 videotapes. This limitation was implemented to maintain reliability

of coding and to prevent “coder drift.”³ Coding reliability begins to falter beyond that number. Initially, all coders worked up to 4 hours a day as coders and then spent the other 4 hours working on other ECLS-B activities, such as field staff payroll, locating respondents, receipting, and computer-assisted data entry (CADE).

Unlike the NCATS coding at 9 months, when inter-lab reliability was maintained between Westat and the NCAST coder at the University of Washington, inter-lab reliability between Westat and the Three Bags Task coder at Teachers College was not maintained beyond the initial training period as had been the case for the Early Head Start Research and Evaluation Project. The Three Bags Task coding staff at Teachers College did not see the need for, and did not have the resources to provide, ongoing inter-lab reliability. Therefore, reliability was maintained within the Westat coding workshop. Intra-lab reliability required that the coding supervisor select a random subsample of ECLS-B Two Bags Task videotapes. The supervisor and an assistant (trained at Teachers College) coded each selected tape independently and then resolved any discrepancies by consensus. When necessary, a third individual (also trained at Teachers College) was brought in to help resolve any particularly difficult coding issues. These selected videotapes (“reliability videotapes”) were then used to establish the reliability of the coding staff.

Two Bags Task coders were then required to code one reliability tape per week, selected at random. Coders were required to code the weekly reliability videotape with a minimum of 85 percent agreement with the consensus reliability coding. If a coder slipped below 85 percent agreement on a weekly reliability videotape, that coder then immediately coded a second reliability videotape. Had there been a case where the second reliability videotape was also below 85 percent agreement, the coder would have been told to cease coding any videotapes from the ECLS-B and would have received remedial training on the identified coding problems. In practice, however, no coder ever slipped below 85 percent agreement on more than one reliability videotape. Agreement rates between the coders and the reliability coding were quite high for the entire coding year and are summarized in table 6-1.

³ Coder drift refers to the change in how information is coded over time by an individual. The coder is said to “drift” from the standard due to such factors as fatigue, forgetting of the rules, failing to detect target information, among others.

Table 6-1. Average reliability (percent agreement) for subscales of the Two Bags Task for the ECLS-B 2-year data collection: 2003–04

	Mean percentage agreement
Two Bags Task scale	
Overall agreement for parent rating scales:	96.5
Parent rating scales	
Sensitivity	97.0
Intrusiveness	98.0
Positive regard	93.0
Cognitive stimulation	94.0
Negative regard	98.0
Detachment	99.0
Overall agreement for child rating scales	94.7
Child rating scales	
Engagement	94.0
Sustained attention	93.0
Negativity	97.0

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.

6.6 Two Bags Performance in the 2-Year National Data Collection

In the Early Head Start Research and Evaluation Project, three of the parent scales were intercorrelated: parental sensitivity, parental stimulation of cognitive development, and the parental positive regard scale. These three variables were combined by Early Head Start to create a Supportiveness composite at 2 years by simply obtaining the mean of the three scales (i.e., the sum of the scores for Parental Sensitivity, Parental Cognitive Stimulation, and Parental Positive Regard divided by 3). For the convenience of researchers, a composite of the average of these three variables also was created for the ECLS-B 2-year data collection. This composite is X2TBSPT.

Table 6-2 presents descriptive statistics for the Two Bags Task variables for the sample as a whole.

Table 6-2. Weighted means and standard deviations for the Two Bags Task rating scales in the ECLS-B 2-year data collection: 2003–04

Two Bags Task scale	Number	Range	Weighted mean	Standard deviation
Parent rating scales				
Sensitivity (C2SENSTV)	7,450	1-7	4.77	0.95
Intrusiveness (C2NTRUSV)	7,450	1-7	1.80	0.54
Positive regard (C2POSRGD)	7,450	1-7	4.26	1.03
Cognitive stimulation (C2COGDEV)	7,450	1-7	4.12	1.08
Negative regard (C2NEGRGD)	7,450	1-7	1.10	0.44
Detachment (C2DETACH)	7,450	1-7	1.05	0.32
Supportiveness (X2TBSPT)	5,600	1-7	4.43	0.86
Child rating scales				
Engagement (C2ENGPRT)	7,450	1-7	4.56	1.14
Sustained attention (C2STNATT)	7,450	1-7	4.47	1.15
Negativity (C2NEGPRT)	7,450	1-7	1.36	0.76

NOTE: The child weight W2C0 was used to produce these statistics. The variable name of the scale is in parentheses. Sample sizes have been rounded to the nearest 50.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.

Although the scores obtained are considered rating scales, they could also be conceptualized as items. Therefore, Cronbach’s alpha was also calculated to investigate whether the scales have a conceptual coherence that would make it feasible to scale them into a single scale. This was done only for the parent scales and did not include the composite X2TBSPT. Cronbach’s alpha was not calculated for the child scales because there are only three reducing the accuracy of Cronbach’s alpha. For all the scales, a value of 1 indicates a low score for that dimension and a 7 indicates a high score. Therefore, it was not necessary to reverse code any of the scores. Cronbach’s alpha for the parent scale was 0.73, indicating that these items have adequate coherence, and the analyst may want to consider creating a single scale comprised of all six parent scales. That said, however, the analyst is cautioned that the rating scales were not designed with that intent and that using them in this way may reduce their usefulness. Instead, the analyst may want to investigate the factor structure of these scores and conduct a factor analysis.

6.7 Correlations of 2-Year Two Bags Task and 9-Month NCATS Scale Scores

Although the Two Bags Task rating scales and the NCATS scale used at 9 months have their differences, the constructs they measure do share some similarities. The NCATS parent scales include items that assess parental sensitivity and responsiveness to the child’s distress, as well as parent fostering

of the child’s socioemotional and cognitive growth. The NCATS child scales include items that assess the child’s responsiveness to the parent and the child’s clarity of communication to the parent. Therefore, both scales measure maternal sensitivity and engagement with the child, as well as the child’s engagement with the parent. The Two Bags Task also shares some characteristics with the child’s responsivity to the parent on the NCATS scale. Therefore, meaningful correlations should be found between the two scales. Table 6-3 summarizes the correlations between the 2-year Two Bags Task rating scales and the 9-month NCATS scales. To obtain these correlations, all cases with missing data were omitted and the child weight, W2C0, was applied.

Table 6-3. Correlation (*r*) of 2-year Two Bags Task rating scales with 9-month NCATS total scale, total parent scale and total child scale, ECLS-B 9-month and 2-year data collections: 2001–02 and 2003–04

Two Bags Task scale	9-month NCATS scales		
	Total scale correlation (<i>r</i>)	Total parent scale correlation (<i>r</i>)	Total child scale correlation (<i>r</i>)
Parent rating scales			
Sensitivity (C2SENSTV)	.19*	.21*	.05*
Intrusiveness (C2NTRUSV)	-.09*	-.11*	-.00
Positive regard (C2POSRGD)	.18*	.21*	.05*
Cognitive stimulation (C2COGDEV)	.17*	.19*	.06*
Negative regard (C2NEGRGD)	-.11*	-.13*	-.02
Detachment (C2DETACH)	-.05*	-.05*	-.04*
Supportiveness (X2TBSPPT)	.22*	.24*	.06*
Child rating scales			
Engagement (C2ENGPRT)	.14*	.15*	.05*
Sustained attention (C2STNATT)	.10*	.10*	.03*
Negativity (C2NEGPRT)	-.07*	-.09*	.01

* $p < .05$

NOTE: The child weight W2C0 was used to obtain these data. $n = 4,900$ (rounded to the nearest 50).

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.

6.8 Two Bags Task Measures in the 2-Year Data Collection

Table 6-4 summarizes the Two Bags Task parent scales that assess characteristics of the primary caregiver’s interaction with the child and table 6-5 summarizes the children’s scales. Both tables present the means and standard deviations for the total sample and by key demographic groups.

Table 6-4. Weighted means (and standard deviations) of the Two Bags Task parent scales by key demographic variables, 2-year data collection: 2003–04

Characteristic	Number	Two Bags Task parent scale variables, weighted means						
		Supportiveness composite (X2TBSPT)	Sensitivity (C2SENSTV)	Intrusiveness (C2NTRUSV)	Positive regard (C2POSRGD)	Cognitive stimulation (C2COGDEV)	Negative regard (C2NEGRGD)	Detachment (C2DETACH)
Total sample	7,450	4.38 (0.88)	4.79 (1.62)	1.19 (1.14)	4.29 (1.83)	4.13 (1.64)	1.12 (0.44)	1.09 (1.90)
Child's race/ethnicity ¹								
White	3,250	4.59 (0.82)	5.00 (0.86)	1.12 (0.61)	4.50 (2.24)	4.35 (1.99)	1.07 (0.33)	1.09 (2.45)
Black	1,200	4.05 (0.83)	4.37 (0.95)	1.41 (0.82)	3.90 (1.08)	3.86 (0.96)	1.32 (0.77)	1.07 (0.40)
Hispanic, race specified	1,000	4.20 (0.90)	4.71 (3.48)	1.27 (2.47)	4.09 (1.04)	3.92 (1.02)	1.10 (0.40)	1.07 (0.32)
Hispanic, no race specified	450	3.95 (0.90)	4.31 (0.99)	1.16 (0.44)	3.85 (0.98)	3.69 (1.02)	1.10 (0.36)	1.13 (0.50)
Asian	700	4.18 (0.87)	4.52 (0.94)	1.24 (0.60)	4.13 (1.03)	3.90 (1.04)	1.15 (0.50)	1.21 (3.11)
Native Hawaiian/Pacific Islander	50	4.25 (0.63)	4.52 (0.66)	1.27 (0.54)	4.57 (1.04)	3.66 (0.71)	1.10 (0.30)	1.04 (0.27)
American Indian/Alaska Native	200	3.93 (0.78)	4.53 (0.85)	1.08 (0.33)	3.76 (1.03)	3.50 (1.01)	1.09 (0.36)	1.07 (0.28)
More than 1 race	600	4.46 (0.84)	4.87 (0.94)	1.13 (0.50)	4.40 (1.07)	4.10 (1.04)	1.09 (0.33)	1.04 (0.29)
Poverty status								
Below poverty threshold	1,600	3.94 (0.85)	4.40 (3.06)	1.37 (2.22)	3.82 (1.07)	3.68 (0.97)	1.23 (0.64)	1.10 (0.45)
At or above poverty threshold	5,850	4.50 (0.85)	4.90 (0.92)	1.14 (0.59)	4.41 (1.96)	4.25 (1.76)	1.08 (0.37)	1.09 (2.12)
Child's sex								
Male	3,800	4.34 (0.88)	4.77 (2.07)	1.22 (1.52)	4.23 (1.04)	4.06 (1.08)	1.14 (0.49)	1.12 (2.55)
Female	3,650	4.43 (0.88)	4.82 (0.95)	1.16 (0.49)	4.34 (2.39)	4.21 (2.07)	1.09 (0.38)	1.06 (0.77)

See notes at end of table.

Table 6-4. Weighted means (and standard deviations) of the Two Bags Task parent scales by key demographic variables, 2-year data collection: 2003–04—Continued

Characteristic	Number	Two Bags Task parent scale variables, weighted means						
		Supportiveness composite (X2TBSPT)	Sensitivity (C2SENSTV)	Intrusiveness (C2NTRUSV)	Positive regard (C2POSRGD)	Cognitive stimulation (C2COGDEV)	Negative regard (C2NEGRGD)	Detachment (C2DETACH)
Birth weight								
Normal	5,500	4.39 (0.88)	4.81 (1.67)	1.17 (0.64)	4.29 (1.88)	4.15 (1.68)	1.11 (0.44)	1.09 (1.97)
Moderately low	1,150	4.29 (0.84)	4.64 (0.92)	1.40 (3.86)	4.20 (1.04)	4.02 (1.01)	1.17 (0.51)	1.06 (0.40)
Very low	800	4.19 (0.86)	4.53 (0.95)	1.32 (0.74)	4.18 (1.05)	3.86 (1.03)	1.18 (0.53)	1.06 (0.37)
Child's age at assessment								
21 months and under	#	4.66 (0.67)	4.58 (0.49)	1.00 (0.00)	4.76 (0.61)	4.65 (1.12)	1.00 (0.00)	1.00 (0.00)
22–23 months	750	4.24 (0.89)	4.58 (0.95)	1.23 (0.61)	4.15 (1.03)	3.98 (1.06)	1.16 (0.50)	1.07 (0.38)
24–25 months	5,750	4.40 (0.88)	4.83 (1.77)	1.17 (0.65)	4.30 (2.00)	4.15 (1.78)	1.11 (0.43)	1.10 (2.15)
26–27 months	750	4.39 (0.90)	4.77 (0.97)	1.27 (3.16)	4.27 (1.06)	4.13 (1.07)	1.11 (0.45)	1.07 (0.40)
28 months and over	150	4.46 (0.78)	4.84 (0.95)	1.14 (0.41)	4.38 (0.88)	4.17 (0.91)	1.11 (0.34)	1.02 (0.16)
Mother's race/ethnicity¹								
White	3,550	4.59 (0.82)	5.00 (0.87)	1.12 (0.61)	4.49 (2.19)	4.33 (1.95)	1.07 (0.33)	1.09 (2.37)
Black	1,200	4.05 (0.83)	4.38 (0.96)	1.41 (0.82)	3.91 (1.08)	3.87 (0.95)	1.33 (0.80)	1.07 (0.40)
Hispanic, race specified	1,200	4.07 (0.91)	4.51 (3.05)	1.24 (2.14)	3.99 (1.02)	3.80 (1.02)	1.09 (0.35)	1.10 (0.42)
Hispanic, no race specified	50	3.92 (0.96)	4.39 (1.02)	1.06 (0.36)	3.65 (0.88)	3.73 (1.17)	1.04 (0.23)	1.14 (0.45)
Asian	850	4.22 (0.85)	4.57 (0.93)	1.21 (0.56)	4.16 (1.02)	3.93 (1.04)	1.13 (0.47)	1.19 (2.81)
Native Hawaiian/Pacific Islander	50	4.36 (0.79)	4.69 (0.92)	1.32 (0.62)	4.38 (1.05)	4.00 (1.03)	1.16 (0.37)	1.04 (0.28)
American Indian/Alaska Native	250	4.09 (0.80)	4.69 (0.87)	1.08 (0.36)	3.96 (1.05)	3.61 (1.00)	1.07 (0.28)	1.05 (0.26)
More than 1 race	200	4.40 (0.86)	4.83 (0.98)	1.14 (0.54)	4.30 (1.03)	4.07 (1.02)	1.13 (0.39)	1.00 (0.07)

See notes at end of table.

Table 6-4. Weighted means (and standard deviations) of the Two Bags Task parent scales by key demographic variables, 2-year data collection: 2003–04—Continued

Characteristic	Number	Two Bags Task parent scale variables, weighted means						
		Supportiveness composite (X2TBSPT)	Sensitivity (C2SENSTV)	Intrusiveness (C2NTRUSV)	Positive regard (C2POSRGD)	Cognitive stimulation (C2COGDEV)	Negative regard (C2NEGRGD)	Detachment (C2DETACH)
Mother's age (in years)								
19 and under	250	3.85 (0.94)	4.24 (1.03)	1.38 (0.80)	3.71 (1.18)	3.60 (0.98)	1.40 (0.93)	1.10 (0.45)
20–29	3,300	4.23 (0.87)	4.63 (0.96)	1.23 (1.50)	4.10 (1.04)	3.95 (1.04)	1.14 (0.49)	1.13 (2.66)
30–39	3,250	4.56 (0.85)	4.95 (0.91)	1.13 (0.46)	4.50 (2.50)	4.31 (1.08)	1.07 (0.33)	1.05 (0.79)
40 and over	550	4.59 (0.85)	5.25 (5.24)	1.14 (1.39)	4.48 (0.97)	4.62 (5.01)	1.07 (0.31)	1.04 (0.31)
Mother's education								
8th grade and under	300	3.67 (0.81)	4.01 (0.95)	1.20 (0.52)	3.61 (0.93)	3.40 (0.84)	1.11 (0.43)	1.22 (0.63)
9–12th grades	1,400	4.00 (0.87)	4.47 (3.07)	1.30 (0.73)	3.89 (1.07)	3.74 (1.00)	1.21 (0.61)	1.26 (4.02)
High school diploma	1,600	4.23 (0.83)	4.62 (0.94)	1.27 (2.11)	4.12 (1.02)	3.94 (0.98)	1.15 (0.51)	1.04 (0.26)
Vocational/technical	150	4.41 (0.77)	4.81 (0.90)	1.13 (0.44)	4.28 (0.84)	4.12 (1.00)	1.04 (0.20)	1.05 (0.29)
Some college	1,850	4.53 (0.82)	4.94 (0.85)	1.14 (0.81)	4.52 (3.18)	4.23 (1.07)	1.09 (0.36)	1.03 (0.19)
Bachelor's degree	1,250	4.78 (0.77)	5.16 (0.80)	1.08 (0.35)	4.60 (0.91)	4.68 (3.25)	1.05 (0.27)	1.04 (1.23)
Graduate school (no degree)	150	4.86 (0.63)	5.27 (0.69)	1.05 (0.28)	4.71 (0.78)	4.59 (1.04)	1.05 (0.22)	1.00 (0.06)
Master's degree	500	4.86 (0.72)	5.27 (0.80)	1.08 (0.38)	4.69 (0.84)	4.63 (1.06)	1.03 (0.23)	1.02 (0.17)
Doctoral/professional degree	200	4.89 (0.77)	5.25 (0.78)	1.01 (0.14)	4.68 (0.82)	4.73 (1.00)	1.02 (0.16)	1.03 (0.34)

Rounds to zero.

¹ Race categories exclude Hispanic origin unless specified.

NOTE: Results were obtained by applying the sampling child weight W2C0. The variable names of the parent scales are in parentheses in column headings. Standard deviations appear in parentheses in the table columns. Detail may not sum to total due to rounding. Sample sizes have been rounded to the nearest 50.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.

Table 6-5. Weighted means (and standard deviations) of the Two Bags Task child scales by key demographic variables, 2-year data collection: 2003–04

Characteristic	Number	Two Bags Task child scale variables, weighted means		
		Engagement (C2ENGPRT)	Sustained attention (C2STNATT)	Negative regard (C2NEGPRT)
Total sample	7,450	4.57 (1.30)	4.51 (2.18)	1.39 (1.95)
Child's race/ethnicity ¹				
White	3,250	4.77 (1.23)	4.72 (2.68)	1.35 (2.15)
Black	1,200	4.32 (1.08)	4.25 (1.06)	1.60 (2.77)
Hispanic, race specified	1,000	4.30 (1.21)	4.24 (1.18)	1.40 (0.84)
Hispanic, no race specified	450	4.14 (1.93)	4.06 (1.86)	1.34 (0.71)
Asian	700	4.32 (1.07)	4.43 (1.09)	1.38 (0.77)
Native Hawaiian/Pacific Islander	50	4.38 (0.98)	4.22 (0.93)	1.27 (0.64)
American Indian/Alaska Native	200	4.21 (0.98)	3.84 (1.20)	1.48 (0.88)
More than 1 race	600	4.71 (1.09)	4.56 (1.07)	1.32 (0.73)
Poverty status				
Below poverty threshold	1,600	4.15 (1.48)	4.09 (1.43)	1.49 (0.89)
At or above poverty threshold	5,850	4.67 (1.22)	4.61 (2.33)	1.37 (2.14)
Child's sex				
Male	3,800	4.44 (1.30)	4.35 (2.06)	1.46 (2.26)
Female	3,650	4.70 (1.28)	4.67 (2.29)	1.32 (1.54)
Birth weight				
Normal	5,500	4.59 (1.31)	4.52 (2.25)	1.39 (2.10)
Moderately low	1,150	4.37 (1.09)	4.31 (1.10)	1.44 (0.84)
Very low	800	4.07 (1.14)	4.00 (1.05)	1.51 (0.81)
Child's age at assessment				
21 months and under	#	4.26 (0.47)	4.14 (0.66)	1.34 (0.48)
22–23 months	750	4.30 (1.71)	4.27 (1.71)	1.38 (0.76)
24–25 months	5,750	4.59 (1.24)	4.52 (2.35)	1.41 (2.17)
26–27 months	750	4.63 (1.17)	4.54 (1.14)	1.33 (0.73)
28 months and over	150	4.83 (1.11)	4.85 (1.16)	1.13 (0.35)

See notes at end of table.

Table 6-5. Weighted means (and standard deviations) of the Two Bags Task child scales by key demographic variables, 2-year data collection: 2003–04—Continued

Characteristic	Number	Two Bags Task child scale variables, weighted means		
		Engagement (C2ENGPRT)	Sustained attention (C2STNATT)	Negative regard (C2NEGPRT)
Mother's race/ethnicity¹				
White	3,550	4.76 (1.22)	4.70 (2.61)	1.35 (2.10)
Black	1,200	4.33 (1.09)	4.26 (1.05)	1.61 (2.76)
Hispanic, race specified	1,200	4.21 (1.45)	4.16 (1.50)	1.38 (0.30)
Hispanic, no race specified	50	4.02 (1.47)	4.12 (1.37)	1.18 (0.47)
Asian	850	4.33 (1.06)	4.43 (1.08)	1.34 (0.72)
Native Hawaiian/Pacific Islander	50	4.53 (0.89)	4.11 (0.98)	1.25 (0.61)
American Indian/Alaska Native	250	4.35 (1.01)	4.07 (1.20)	1.40 (0.84)
More than 1 race	200	4.71 (1.12)	4.50 (1.13)	1.39 (0.82)
Mother's age (in years)				
19 and under	250	4.29 (1.14)	4.17 (1.13)	1.44 (0.91)
20–29	3,300	4.47 (1.32)	4.42 (2.34)	1.43 (1.62)
30–39	3,250	4.70 (1.30)	4.62 (2.19)	1.35 (2.41)
40 and over	550	4.57 (1.08)	4.51 (1.11)	1.31 (0.60)
Mother's education				
8th grade and under	300	3.78 (1.19)	3.78 (1.11)	1.42 (0.77)
9–12th grades	1,400	4.26 (1.16)	4.30 (3.15)	1.47 (0.86)
High school diploma	1,600	4.46 (1.10)	4.36 (1.10)	1.39 (0.80)
Vocational/technical	150	4.63 (3.17)	4.56 (3.15)	1.23 (0.63)
Some college	1,850	4.68 (1.08)	4.57 (1.11)	1.39 (2.07)
Bachelor's degree	1,250	4.89 (1.54)	4.85 (3.25)	1.40 (3.81)
Graduate school (no degree)	150	4.96 (1.05)	4.74 (1.25)	1.32 (0.83)
Master's degree	500	5.03 (1.01)	4.89 (1.11)	1.20 (0.67)
Doctoral/professional degree	200	4.97 (0.94)	4.99 (1.08)	1.21 (0.51)

Rounds to zero.

¹ Race categories exclude Hispanic origin unless specified.

NOTE: Results were obtained by applying the sampling child weight W2C0. Standard deviations appear in parentheses. Detail may not sum to total due to rounding. Sample sizes have been rounded to the nearest 50.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), 2-year data collection, 2003–04.