

Development of a short form of the Bayley Scales of Infant Development, Third Edition to Assess Children's Developmental Status in the National Children's Study

Louise O'Donnell, University of Texas Health Science Center-San Antonio
Carol Andreassen, Kathryn Bojczyk, and Philip Fletcher, Westat
Marsha Gerdes and Sofia Baglivo, Children's Hospital of Philadelphia
Susan Spieker and Thomas Burbacher, University of Washington

The NCS requires a psychometrically robust measure of children's developmental baseline status in order to compare it to other outcome measures, measure cognitive outcomes of at-risk children with negative exposure histories, and compare NCS results with other studies of early child development. The Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-3) is the gold standard measure of choice and has been used in large-scale studies (e.g., ECLS-B, SECCYD). The Bayley-3 has 3 subscales: Cognitive Scale, Motor Scale and Language Scale. Our goal was to develop age-specific short forms for the 6-, 12-, 18-, 24- and 36-month visits that would be as psychometrically rigorous as the full Bayley 3, would streamline administration, reduce burden to participants and data collectors, and measure children's performance across the entire ability range.

Methods: With the publisher's permission, the standardization dataset (N=1,700) was obtained. Item Response Theory analysis (IRT) was conducted to select an appropriate subset of items. Item selection was also guided by administrative ease, objectivity of scoring, and minimal use of stimulus materials. Administration and scoring instructions were reformulated for maximal field efficiency. Data collection from a representative sample of children is ongoing or planned at 4 sites (University of Texas Health Science Center at San Antonio, Children's Hospital of Philadelphia, University of Washington, and Westat), with each site collecting Bayley-3 short form data on approximately 250-300 children, for a total of 1,000 children across age groups.

Results: IRT is a powerful psychometric tool used in test construction and analysis to generate item parameters used in scaling, scoring, and item selection. Item difficulty parameters, discrimination parameters, and Item Characteristic Curves (ICC) were used to select short form items. Streamlined and uniform basal and ceiling rules were also developed using IRT to simplify administration: fewer than 3 items correct requires administration of "easier" items; fewer than 3 items incorrect requires administration of "harder" items.

Discussion: Each short form has an IRT forecast reliability of .80 or greater. When pilot data collection is completed, IRT analysis will be conducted to establish the psychometric properties of each short form.