Bayley scales of Infant and Toddler Development
Third edition

Carol Andrew, EdD, OTR
Assistant Professor of Pediatrics
Dartmouth Hitchcock Medical Center
Lebanon, New Hampshire, USA
Revision goals

► Update the normative data
► Strengthen the psychometric quality
► Simplify administrative procedures
► Update stimulus materials
► Develop “all-domain” assessment capability
Why revise?

- Improve quality and utility of the instrument
- Reflects new research on child development
- Demographic shifts over the past 15 years or so since BSID II was normed
Update norms

► Normative data collected in USA from January 2004-October 2004

► Sample was stratified on key demographic variables
  - Age, sex, race/ethnicity, parent education level, geographic location
  - Children with disabilities or delays excluded from standardization

► Special data collected on children with specific clinical diagnoses
Strengthen Psychometric qualities

- New items added to strengthen and expand content coverage
- Reliability evaluated for normative and clinical groups
- Validity gathered for internal structures and in comparison to other measures such as WPPSI III and PLS-4.
- Subtest discrepancy analysis available
Simplify administrative procedures

► No “item sets”
► Reverse and discontinue criteria same for all subtests
► Items clearly marked as part of a series
► More understandable directions
► More sample “acceptable” responses
Update stimulus materials

- Materials are more colorful, interesting, artwork updated
- Picture book and story book sturdy cardboard
- Stimulus book includes easel and directions on tester’s side of the book
- Quality of toys improved
Develop all domain capability

- Three \textit{administered} scales, five subtests
  - Cognitive scale
  - Language scale
    - Receptive
    - expressive
  - Motor Scale
    - Fine motor
    - Gross motor

- Required for Individuals with Disabilities Education Act (IDEA)
Develop all domain capacity...

► Two scales derived from caregiver questionnaires
  - Social emotional
  - Adaptive Behavior
Item Bias examined

- Experts in cross-cultural research and/or child development conducted formal reviews of items twice during development.

- Problematic items were identified and deleted (13 from cognitive scale, 4 from receptive language scale, 3 from expressive language scale, 5 from motor scale).
Evidence for Reliability

- Internal consistency of subtest
  - Average reliability (Fishers z transformation) across age groups
    - Cognitive scale: .91
    - Expressive communication: .91
    - Receptive communication: .87
    - Gross Motor: .91
    - Fine Motor: .86
Special Group Reliability

- 668 children from special groups
  - Down syndrome
  - At risk for delay
  - Language impairment
  - Cerebral palsy
  - Pervasive developmental disorder
  - Alcohol exposure
  - Prematurity

- Cognitive .96
- Receptive language .95
- Expressive language .96
- Fine motor .95
- Gross Motor .98
Evidence of Validity

- Comprehensive literature searches and expert reviews conducted to ensure content validity (chapters 2-3 of Technical Manual)
- Social emotional scale and Adaptive Behavior scales derived from existing scales with extensive research supporting their validity
Correlations with other measures

- BSID II
- Wechsler Preschool Primary Scale of Intelligence III
- Preschool Language Scale–4
- Peabody Developmental Motor Scale – 2
- Adaptive Behavioral Assessment System II

Details available in technical manual
Special group studies

- Pervasive developmental disorder/autistic spectrum disorder  N=70
- Down syndrome  N=90
- Language impairment  N=94
- Small for gestational age  N=44
- Fetal alcohol exposure  N=48
- Cerebral palsy  N=73
- Birth asphyxia  N=43
- Premature  N=85
Special issues in using BSID-III internationally

- Standardized in English: language items may not translate well
- Many more items – better validity but takes longer (estimate 90 minutes for children over about 18 months of age)
- Questionnaires for caregivers very complete but very laborious for parents or other caregivers to do
Still have all the issues of infant assessment:

- Prognostication poor unless severe delay (<68)
- Child and parent comfort with the situation essential to appropriate sampling of behavior
- Only measures current development during the time observed – does not provide diagnosis or treatment recommendations
Issues for Premature Followup

► BSID II is quicker to administer and score (60 minutes instead of 90 minutes)
► In the US, already using BSID II for Vermont-Oxford data collection and will continue
► BSID II does not measure quality of movement so Motor Scale (PDI) is not used for “neuromotor” status
► Vast majority of our premature are delayed on PDI through age 2 but neurologically “normal”
When and why to use BSID III

► If beginning a research project which requires a very careful evaluation of all aspects of a child’s performance, where time is no problem
► For a complete “one-stop” evaluation which suggests need for further assessment in specific area (eg, speech pathology referral, occupational therapy referral, etc)
► For early intervention “team assessments”
► When following a child over time to make a diagnosis or to document change in status (such as neurometabolic conditions with potential for regression)
Subtests can stand alone

► Can be utilized for identification of specific issues such as language delay or motor delay or cognitive delay in a given population

► Each subtest takes perhaps 15 minutes to administer at the most, depending on the age of the child
Bayley Screening Test

► Multiple subtests
► Abbreviated format
  ▪ Cognitive 33 items instead of 91 items
  ▪ Receptive language 24 instead of 49 items
  ▪ Expressive language 24 instead of 48 items
  ▪ Fine motor 27 instead of 66 items
  ▪ Gross motor 28 instead of 72 items

*Results are not numeric: instead have “at risk”, “emerging”, “competent” designations
Summary of BSID III in comparison with BSID II

**Benefits:**
- More complete evaluation
- Clearer directions
- Better materials
- Subtests can stand alone
- Parent report separate and complete
- PDA scoring available

**Problems:**
- Takes much longer to administer
- If research already underway with BSID II, can’t switch
- Language items may present difficulty for non-English language structures