Evidence Summary for Pediatric Rehabilitation Professionals

Outcome Measures: The Bayley Scales of Infant Development, 3rd Edition (BSID-III)

1. Summary

Type: Norm-referenced
Purpose: Discrimination, planning
Population: Children at risk for developmental delay
Age: 0 – 42 months
Time to Complete: 50 - 90 minutes
Equipment Needed: BSID-III kit, table, chair, floor space for motor items

2. Overview

The BSID-III\(^1\) (2006) is a revision of the BSID-II\(^2\) (1993). The primary purpose of the BSID-III is to identify children with developmental delay and to provide information for planning intervention.\(^1\) The goals of this most recent revision were to:

a. Update the normative data by developing new norms that better represent societal change in development and allow increased precision in the infant period in which children are rapidly developing.
b. Develop 5 distinct scales which meet the requirements set by the United States Individual with Disabilities Education Improvement Act of 2004.
c. Strengthen the psychometric quality by:
   i. Extending floors and ceilings to better measure significant developmental delay as well as advanced development
   ii. Improving evidence of reliability and validity
   iii. Re-examining item bias to remove items with potential bias and datedness, and ensure content relevance and clinical utility
   iv. Simplify administration procedures by removing items sets and including reversal and discontinue rules for all subtests as well as simplifying instructions and reorganizing the manual
   v. Update item administration by making the instructions more play based, reducing the effect of receptive and expressive language on cognitive items, and allowing caregiver involvement providing administration procedures are followed
   vi. Update stimulus materials to allow selection of materials that appeal to the child and make materials more appealing
   vii. Maintain basic qualities of the Bayley Scales

The BSID-III is composed of five subscales:

a. Cognitive subscale
b. Language scale
   i. Receptive language subscale
   ii. Expressive language subscale
c. Motor scale

i. Fine motor subscale

ii. Gross motor subscale

The BSID-III is intended to be administered by individuals who have training and experience in the administration and interpretation of comprehensive developmental assessments.1

Administration of each scale is started at a pre-determined item, based on the child’s age.1 A child must achieve a score of 1 on the first three consecutive items at the pre-determined start point to achieve the basal score. If this is not the case, administration begins at the previous start point ("reversal rule"). The reversal rule continues to apply until the child has achieved the first three consecutive items at the determined start point. To complete testing and achieve the ceiling, a child must score 0 on five consecutive items. After having received these five consecutive 0 scores, no further items are administered ("discontinue rule").1

When scoring, each of the five subscales is given a raw score based on the number of items the child has achieved in addition to the number of unadministered items preceding the basal.1 Higher scores indicate more mature development. From these raw scores, scaled scores can be calculated for the cognitive scale and the two combined language scales and motor scales. These scores can then be used to determine composite scores, percentile ranks and confidence intervals, developmental age equivalents, and growth scores.1,2

In addition to the standard BSID-III, a parent/caregiver questionnaire has been developed and includes social-emotional and adaptive behavior scales.1,2 The social-emotional scale consists of the Greenspan Social-Emotional Scale3 and the adaptive behavior scale consists of the Adaptive Behavior Assessment System, 2nd edition4 (ABAS-2).1

3. Standardization Sample

The BSID-III was standardized on a normative sample of 1700 children between the ages of 16 days and 43 months living in the United States in 2004.1 Stratification was based on age, gender, parent education level, ethnic background, and geographical area.1

Norms for the social-emotional and adaptive behavior scales followed the same stratification pattern but were derived from smaller groups (456 and 1350 children respectively).1

4. Measurement Properties

a. Reliability

i. BSID-III

The BSID-III has established reliability. Internal consistency was assessed using a split-half reliability method and shows reliability coefficients for the subscales and composite scores that range from 0.86 to 0.93.1 Reliability coefficients for the special groups assessed are similar or higher than those of the normative sample indicating that the BSID-III is equally reliable for children with clinical diagnoses or risk factors as for the general population.1

Test-retest reliability of the cognitive, language, and motor scales was assessed by testing 197 children twice.
Outcome Measures: The BSID-III

over 2 to 15 days. Correlation scores varied depending on the children’s ages and ranged from 0.67 to 0.94 for the different subtests. Changes in scores over this very short period of time is attributed to maturation and practice effects. Test-retest reliability for the adaptive behaviour scale was calculated by asking 207 parents to rate their child twice over 2 days to 5 weeks. Reliability coefficients ranged from 0.71 to 0.92. Standard error of measurement and confidence intervals were also calculated.

ii. BSID-III Social Emotional Scale

Internal consistency of the social emotional was also examined. Scores from the Greenspan social-emotional growth chart, which makes up the social-emotional scale, indicate strong internal consistency with coefficients ranging from 0.76 to 0.94.

iii. BSID-III Adaptive Behavior Scale

Internal consistency scores for the ABAS-II, which forms the adaptive behavior scale show coefficients ranging from 0.79 to 0.98 for the normative sample and 0.75 to 0.99 for the special populations groups. Inter-rater reliability on the adaptive behavior scale was assessed by having two parents/caregivers rate the same child. Results on 56 children show reliability coefficients ranging from 0.59 to 0.82.

b. Validity

i. BSID-III

The BSID-III also has established convergent and divergent validity, as illustrated by the correlations between the BSID-III and other relevant instruments. Resultant correlations were consistent in direction and magnitude with the author’s expectations.

The relationship between the BSID-III and BSID-II was examined. As expected the language composite score of the BSID-III showed greater correlation with the BSID-II mental index \((r = 0.71)\) than with the motor index \((r = 0.47)\). As well, the BSID-III motor composite score showed greater correlation with the BSID-II motor index \((r = 0.60)\) than with the mental index \((r = 0.44)\). However, in clinical practice, previous scores from the BSID-II cannot be compared to scores on the BSID-III since the scores are derived in different manners.

Correlations with the Wechsler Preschool and Primary Scale of Intelligence, 3rd edition (WPPSI-III), a test measuring intellectual functioning, were between 0.52 and 0.83. Correlations with the Preschool Language Scale, 4th edition (PLS-4), a measure of receptive and expressive language, were between 0.50 and 0.71. Lastly, correlations between the Peabody Developmental Motor Scales, 2nd edition (PDMS-2), a test examining motor skills, were between 0.55 and 0.59.

The validity of the BSID-III in children with specific conditions or risk factors was also examined. These “special groups” included: Down syndrome, pervasive developmental disorder, cerebral palsy, specific or suspected language impairment, asphyxiation at birth, prenatal alcohol exposure, small for gestational age, premature or low-birth weight, and other children at-risk for developmental delay. Results
indicate that the BSID-III is sensitive to differences in performance of typical children and children at risk for developmental delay.1

ii. BSID-III Social Emotional Scale

No specific information is provided regarding the validity of the BSID-III social emotional scale.

iii. BSID-III Adaptive Behavior Scale

Correlations of the BSID-III adaptive behavior scale with the Vineland Adaptive Behavior Scale – Interview Edition8 (VABS-IE), a measure that assesses adaptive behavior, show moderate correlations between both outcome measures (0.49 to 0.68).1 Correlations of the BSID-III adaptive behavior scale with the Scale of Independent Behavior – Revised: Early Development Form9 (SIB-R), a brief measure of adaptive behavior, are low (0.18).1

5. Further Considerations

Due to the BSID-III being only recently available, the only literature available regarding its use or psychometric properties is found in the test’s manuals. With time, further information regarding this edition’s strengths and limitations should become available through peer-reviewed publications.

References


This evidence summary is one part of a series on pediatric rehabilitation outcomes measures. Other summaries in this series include:

- Outcome Measures: A Primer
- Outcome Measures: The Alberta Infant Motor Scale (AIMS)
- Outcome Measures: The Bruininks-Oseretsky Test of Motor Performance, 2nd Ed. (BOT-2)
- Outcome Measures: The Developmental Test of Visual-Perception, 2nd Ed. (DTVP-2)
- Outcome Measures: The Gross Motor Function Measures (GMFM)
- Outcome Measures: The Movement Assessment Battery for Children, 2nd Ed. (MABC-2)
- Outcome Measures: The Peabody Developmental Motor Scales, 2nd Ed. (PDMS-2)
- Outcome Measures: The Sensory Profile (SP)