

Gagnon, S. G., & Nagle, R. J. (2000). Comparison of the revised and original versions of the Bayley Scales of Infant Development. *School Psychology International*, 21(3): 293-305 (Aug 2000). Published by SAGE (ISSN: 0143-0343). doi:10.1177/0143034300213006

Comparison of the Revised and Original Versions of the Bayley Scales of Infant Development

Sandra G. Gagnon and Richard J. Nagle

ABSTRACT

Differences between Mental Scale scores on the revised Bayley Scales of Infant Development (BSID-II) and the original test (BSID) were evaluated. Both versions of the measure were administered to 32 subjects, ranging in age from 3 to 28 months, in counterbalanced order. Inter-test correlations indicated that the two versions of the test were significantly correlated ($r = .76, p < 0.05$). Significant differences between scores on the BSID-II ($M = 105.97$) and the original BSID ($M = 124.38$) were observed ($t(31) = 8.41, p < 0.05$). Children below the age of 12 months displayed slightly larger discrepancies than children above 12 months of age. This discrepancy did, however, appear to increase significantly with the developmental levels of the children, with subjects scoring above one standard deviation on the BSID displaying larger discrepancies than those at or below the normal range. These findings are important for psychologists, early interventionists and other professionals involved in the assessment of infants and toddlers as they interpret scores on the new Bayley Scales. This information is especially pertinent for children whose performance is evaluated longitudinally, since they will likely score lower on the new test than they did on the original Bayley Scales. The lower scores may lead to subsequent increases in the number of children who qualify for early intervention and special education services. Therefore, these findings are also important for those involved in the development and evaluation of intervention programs for young children. The findings are considered in terms of their applicability across cultures.

The Bayley Scales of Infant Development (BSID; Bayley, 1969) are well established as one of the most useful instruments available in the assessment of infants (Burns et al., 1992). The Bayley Scales of Infant Development, Second Edition (BSID-II; Bayley, 1993) were introduced to the psychological community in 1993 as a revision of the original Bayley Scales. The aim of the revision was to improve the quality and utility of the BSID, while preserving the original content and intent of the test. Although much of the original content of the BSID was maintained, new items were added to improve the content coverage and ease and accuracy of administration.

The original Bayley Scales have proven to be a sound instrument and have been frequently utilized in research studies with a variety of clinical samples, including children with motor and speech delays (Dilworth and French, 1990), HIV infection and prenatal drug exposure (Mellins et al., 1994) and children considered to be biologically high-risk (Aylward et al., 1995; Crowe et al., 1987; Gerken et al., 1994; Siegel et al., 1995). Numerous authors have provided empirical evidence for the validity and reliability of the BSID (Cook et al., 1989; Snyder and Sheehan, 1992), thus establishing the psychometric integrity of the scales.

The need for revisions and updated normative data prompted the development of the BSID-II. Empirical evidence for the need to revise the BSID was demonstrated by Campbell and colleagues (1986). In their study of 436 twelve-month-old infants of low socioeconomic status these researchers found that on average, subjects scored 7 to 14 points above the normative sample mean Mental Development Index score of 100. The elevated scores found in their sample led these researchers to conclude that the norms for the original Bayley were outdated and were no longer reflective of the current population of children.

According to Bayley (1993), there were seven major goals in the revision of the BSID: to update the normative data, to extend the age range, to improve content coverage, to update the stimulus materials, to improve the psychometric quality and clinical utility of the Scales and to preserve the basic qualities of the original Bayley Scales. To accomplish these goals, Bayley surveyed BSID users to identify strengths and weaknesses of the original scales. Among the strengths were the flexible order of administration of test items, the interesting nature of the tasks and the incorporation of the behaviour rating scale. The weaknesses identified were the layout of the kit, the lack of clarity in the directions and the poor durability of the materials. In revamping the test, Bayley attempted to improve on these weak areas while maintaining the strengths. Approximately three-quarters of the original Mental Scale items were kept in the BSID-II.

Although the BSID-II has been in use since 1993, the existing literature on this measure is limited. The majority of available information on the psychometric properties of the revision is found in the test manual (Bayley, 1993). Evaluation of the reliability of the Mental Scale with the

standardization sample revealed reliability coefficients ranging from 0.78 at 10 months to 0.93 at 27 months. These coefficients, calculated with coefficient alpha and based on sample sizes of 100 for each age group, are consistent with those obtained on the original BSID and indicate high reliability. The test-retest stability of the Mental Scale was assessed in a study of 175 children in four age groups drawn from the standardization sample. Intervals of 1 to 16 days took place between testing. To determine if reliability varied as a function of age, the 1 and 12 month old children and the 24 and 36 month old children were grouped together. Results indicated that test scores were highly stable over time and across age groups ($r = 0.83$ for 1 and 12 months; $r = 0.91$ for 24- and 36-months). Inter-scoring agreement was evaluated by correlating the scores produced by an evaluator and a rater who simultaneously recorded scores in a sample of 51 children. Inter-scoring reliability was found to be high ($r = 0.96$).

Comparison studies of the BSID-II are also described in the manual. To establish the relationship between the BSID-II and the original BSID, both versions of the scales were administered in counterbalanced order to 200 children ranging in age from 1 to 42 months ($M = 15.51$, $SD = 10.13$) (Bayley, 1993). A moderate correlation ($r = 0.62$) was found between the MDI scores on the two tests, as well as a mean difference of 12 points, with the BSID-II scores being lower. These findings indicate substantial overlap in the content of the original and revised versions of the Bayley Scales, although Nellis and Gridley (1994) emphasize that this moderate correlation reflects the notable differences between the two tests and highlights the importance of avoiding generalizing too much about the BSID-II from the BSID. This issue will be addressed further in relation to the findings of the current study. Strong correlations between MDI scores and the General Cognitive Index Score on the McCarthy Scales of Children's Abilities ($r = 0.79$) and the Full Scale IQ of the Wechsler Preschool and Primary Scale of Intelligence - Revised ($r = 0.72$) provide additional evidence for the concurrent validity of the BSID-II.

The results of a comparison study conducted by Goldstein and colleagues (1995) provided support for the clinical validity of the revised Bayley Scales. In this study, both versions of the test were administered to 49 high-risk, premature infants ranging in age from 12 to 22 months. Findings indicated that the mean MDI score was within 4-5 points of the score reported in the BSID-II manual with a similar clinical population. This study also reported very high correlations between MDI's on the BSID and BSID-II, with scores on the new test approximately 8 points below those on the old test. These findings suggest that the constructs measured by the two tests are very similar.

Although the Bayley Scales were standardized on American children, the measure has been used in studies cross-culturally. Whereas the normative data for the BSID may not be applicable to infants from other cultures, developments in other parts of the world indicate that modifications

to the original measure may result in appropriate use of the BSID in other cultures. A Dutch translation of the BSID was developed and administered to 1283 Dutch infants, resulting in the development of Dutch norms for the Mental and Motor scales, as well as for the Infant Behavior Record (van der Meulen and Smrkovsky, 1985). In India, the original Bayley Scales were used in the development of an infant scale of mental development for children from 1 to 30 months, which was found to effectively discriminate between normal and delayed infants (Jayashankarappa, 1986). The original version of the Bayley Scales was used in a study of undernourished Kenyan infants which assessed early predictors of cognitive development (Whaley et al., 1998). The modifications made to the BSID made it more appropriate for the culture and testing conditions. A recent study conducted in Australia compared both versions of the Bayley Scales in a sample of 97 healthy infants ranging in age from 18 to 27 months, revealing 4 to 35 point gains in MDI scores (Tasbihsazan et al., 1997). Thus, it becomes evident that the Bayley Scales can be used appropriately in other cultures. The comparability of the two versions of the test should be studied in the American population before the new version is used in other cultures, since the test was normed on this population. Once this has been demonstrated, new studies should emerge which compare the old and new versions of the measure in different populations.

As with any revised test, it is necessary to examine the relationship between the present and previous versions for both empirical and clinical reasons. The purpose of the present study was to compare Mental Scale scores on the BSID with corresponding scores on the BSID-II among healthy, nonreferred infants. Research by Flynn (1984; 1987) suggests that tests which have been in use for some time tend to produce inflated scores and that with revisions of such tests, mean differences increase with the length of time between standardizations. Bracken (1988) describes how scores on an original and revised version of a test are significantly different, with the magnitude of the difference being dependent on the norming dates. Nellis and Gridley (1994) explain that declines in scores can be expected when a revision of a measure is published many years after the original version. The BSID-II manual explains that these declines should be expected, based on changes in the content, scoring criteria, item directions and floor and ceiling properties (Bayley, 1993). With this information in mind, it was expected that the present sample would demonstrate a decline in scores from the BSID to the BSID-II.

METHODS

Participants

Thirty-two children participated in the study with parental permission. Subjects all resided in the Southeastern United States, specifically in either Georgia or South Carolina. The 11 male and 21 female subjects ranged in age from 3 months to 28 months ($M = 12.56$ months, $SD = 7.31$). All subjects were Caucasian and were drawn from day care centres

and play groups serving middle to upper-middle class families. All subjects came from intact families. Thirty-nine percent of subjects' parents held college degrees, while 27 percent held graduate or professional degrees.

Instruments

The Mental Scales of the Bayley Scales of Infant Development (BSID) (Bayley, 1969) and the Bayley Scales of Infant Development, Second Edition (BSID-II) (Bayley, 1993) were the instruments used in this study. A child's performance on these measures is an indication of the child's developmental status at a specific age. Administration of the Mental Scales for each tests yields a raw score which represents the total number of items passed. The raw score is converted into the Mental Development Index (MDI) score by referring to the norms tables for the child's age derived by Bayley (1969; 1993).

The reliability and validity of the BSID is described in the test manual. The split-half reliability coefficients, based on raw scores across the age ranges from 2 to 30 months, ranged from 0.81 to 0.93, indicating high reliability. Tester-observer agreement was found to be 89.4 percent (SD = 7.1; $N = 90$) and test-retest consistency was found to be 76.4 percent (SC = 13.7; $N = 28$). The research outlined in the test manual also provides support for the content, construct, predictive and discriminant validity of the BSID.

The BSID-II test manual reports reliability coefficients for the Mental Scale that range from 0.78 at 10 months ($N = 100$) to 0.93 at 27 months ($N = 100$), with a mean reliability coefficient of 0.88 across the age ranges from 1 to 42 months. The test-retest stability coefficient for the Mental Scale was reported to be 0.83, with mean scores reported at the first testing ($M = 100.21$, $SD = 14.77$) being consistent with scores at the second testing ($M = 103.26$, $SD = 16.72$). The time intervals between the two testings ranged from 1 to 16 days (median retest interval = 4 days). A high level of interscorer agreement is reported for the Mental Scale ($r = 0.96$, $N = 51$). Support for the validity of the BSID-II is largely based on the validity research reported for the BSID. Given the relationship between the two measures, it appears that the BSID-II is a valid measure.

Procedure

The BSID and BSID-II were administered to subjects in counterbalanced order. Testing was completed in subjects' homes or in their day care centres and was administered by a doctoral student in school psychology with internship training in a paediatric setting. In most cases, subjects' mothers or a familiar adult were present during testing.

RESULTS

Correlations between the two versions of the test were high ($r = 0.76, p < 0.05$) and findings revealed a significant mean difference ($t(31) = 8.41, p < 0.05$) in Mental Development Index (MDI) scores between the BSID ($M = 124.38, SD = 19.09$) and the BSID-II ($M = 105.97, SD = 14.48$). When the data were examined by age groups, significant differences between MDI scores were revealed for children between 1 and 12 months of age and for children over 12 months. Discrepancies between tests were approximately 5 points larger for children in the 1 to 12 month age range ($M = 20.2, SD = 2.82$) than for children over 12 months ($M = 15.42, SD = 3.43$), suggesting that the differences are more likely among lower age levels. The data were also examined according to developmental levels. Results indicated that as children's developmental levels increase, as evidenced by their MDI scores on the BSID, the discrepancies between scores increase. As can be seen in Table 1, subjects whose MDI scores were between 116 and 130 displayed significant differences between scores on the two tests ($t(11) = 8.75, p < 0.05$) as did subjects whose scores were greater than 130 ($t(13) = 8.63, p < 0.05$). An additional analysis described in the BSID-II manual (Table 6.10, p. 216) provided ranges of expected MDI scores for designated scores on the BSID. These ranges tended to be narrow for scores falling in the middle of the distribution and wider as scores moved toward the upper and lower extremes. In the current sample, 5 of the 32 subjects, or 15.6 percent had scores that fell within these predicted ranges.

Table 1 *Discrepancy Scores by Development Level Using BSID Scores*

<i>Developmental level</i>	<i>MDI score</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Below average	<85	-0.50	4.5	2
Within normal limits	85-115	4.00	1.63	6
Above normal limits	116-130	22.18	2.54	11
Accelerated performance	>130	24.76	2.87	13

DISCUSSION

These findings raise important issues regarding the interpretation of scores on the BSID-II and highlight the importance of avoiding inferring too much about the BSID-II from the BSID. Existing research on the revised test indicates that the reliability and validity of the new measure are good. The current study found a slightly stronger correlation between the BSID and BSID-II ($r = 0.76$) than is cited in the BSID-II test manual ($r = 0.62$). The BSID reported split-half reliability coefficients ranging from 0.81 to 0.93 for the mental scale, along with 89.4 percent testerobserver agreement and 76.4 percent agreement for test-retest reliability.

Although the original Bayley manual did not report a test-retest reliability coefficient, the information presented here indicates that the measures are relatively consistent with one another. Results also suggest that children may be expected to score significantly lower on the revised test than on the original BSID. Although Bayley (1993) cited a mean difference of 12 points in MDI scores, children in the present sample displayed discrepancies of approximately 18 points (range = 0 to 38 points) between scores on the two tests. These findings are virtually identical with results from a study of Australian children, in which differences of 18 points (range = 4 to 35 points) were found (Tasbihsazan et al., 1997). Similarities and differences between the Australian study and the present investigation will be presented.

Although the findings are similar, there are both cultural and age differences between these two studies. The average age of subjects in the present study was approximately 12 months, while the average age in the Australian sample was approximately 24 months. Since the majority of the item changes in the revised version of the BSID pertained to the 18 to 27 month age range and the test is strongly dependent on vocabulary items at this age range, Tasbihsazan et al. (1997) chose to focus on this particular age group. It is also important to consider the effects of language when making cross-cultural comparisons. Although the Australian study involved a different culture, this culture and American culture generally share the same language. However, the Australian sample also included a subsample of non-English speaking children. These children scored approximately 16 points lower on the BSID-II, a difference that is consistent with findings with the English speaking children. These similar disparities suggest that the score differences are consistent across cultures and across languages. This trend in findings indicates the need to compare the measure across cultures and languages.

There are several possible explanations for the large magnitude of discrepancies obtained in the present sample. It has been noted in the literature that test scores tend to inflate over time as the result of time between publications of the different versions of the test (Flynn, 1984, 1987; Bracken, 1988; Nellis and Gridley, 1994). Bracken (1988) describes the relationship between test scores and publication dates, explaining that tests normed years ago typically produce higher scores than tests that are normed more recently. According to Bracken, the magnitude of the discrepancy between scores is in direct proportion to the amount of time between dates of publication. This information is important to consider in relation to the present findings, considering the 26-year time span between publications of the BSID and BSID-II. Bayley (1993) also provides explanations for the moderate correlations between the two tests. The many changes and additions to the content, directions for administration and scoring criteria contribute to variability in children's performance on the two measures. In addition, Bayley explains that the floor and ceiling are greater in the BSID-II than in the BSID and this also affects the relationship between the tests. Bracken (1988) describes how floor and ceiling effects can produce

differences in scores on similar tests. Tests with inadequate floors or ceilings do not allow the most accurate distinction of ability levels. The floor and ceiling have been expanded and improved in the BSID-II and this must have some impact on the discrepancies in scores.

Campbell et al. (1986) have suggested that the declines in scores on the BSID-II for children from low socioeconomic backgrounds may be due to the increased availability in the 1970's of federally funded programs which targeted infant growth and development. In their sample of low socioeconomic status children, these authors suggested that these programs resulted in increased access to family planning services, health care and nutritional information, prevention techniques and early intervention services, which subsequently led to increases in the developmental levels of children in the United States. This hypothesis provides another contributing factor for children outperforming the original BSID standardization sample.

In addition to the overall discrepancies in scores, the current study also found differences that emerged among age and developmental levels. Children between 1 and 12 months of age displayed slightly larger discrepancies between test scores than children over 12 months. Results indicate that as the developmental levels of children increase, the discrepancies between scores on the BSID and BSID-II increase. These possible regression effects are addressed in the BSID-II manual (Bayley, 1993), which presents ranges of expected scores on the BSID-II as predicted by scores on the BSID. Scores that fall closer to the mean index score of 100 have narrower ranges of expected scores, or bands of error, than those that fall in the extreme upper and lower scores. In the present sample, 15.6 percent of subjects scored within these predicted ranges. According to the test manual, predictions of BSID-II scores become increasingly less accurate as scores on the BSID move further from the mean and the standard estimate of error bands become wider. In other words, the further from the mean the scores are, the less symmetrical the confidence intervals become and the greater the tendency is for scores to regress to the mean.

It seems important to investigate the factors that may have contributed to the large percentage of subjects scoring above the average range. The influence of the subjects' socioeconomic levels may also have contributed to their levels of performance. Results of a study by Rubin and Balow (1979) revealed a significant correlation between socioeconomic status and performance on the BSID. Specifically, in their sample, children of low socioeconomic status and low BSID scores had the lowest scores on future measures of intelligence. Those children with high socioeconomic status and low BSID scores eventually performed in the average range on intelligence tests. In a longitudinal study of social and environmental influences on the cognitive development of low birth weight infants, Escalona (1984) found socioeconomic status to be a significant contributing factor to cognitive development. Thus, the effects of socioeconomic status on performance on cognitive measures becomes evident. The present sample consisted of children from middle to upper-middle

class families whose parents had relatively high levels of education. This sample characteristic likely contributed to the high scores obtained on the BSID by the majority of subjects.

Cultural differences are also important to consider when attempting to generalize findings from the present study. This Caucasian sample may not be representative of children and families of other ethnic backgrounds or nationalities. In generalizing, it is important to consider factors which may affect parenting in other countries. Aspects of the culture affect parental goals and resulting parenting practices. For example, in the United States, learning object names has cultural importance. Consequently, mothers focus on teaching their young children names of objects so that they are competent within their culture (Greenfield and Suzuki, 1998). Therefore, American children may be better able to respond appropriately to naming items on tests like the Bayley Scales. Differences in the speech parents use with their children in different cultures have been found to contribute to differences in the children's use of nouns and verbs (Tardif et al., 1997). Thus, children from different cultures may perform differently on language items on the Bayley Scales. Sensitivity to these differences will affect psychologists' expectations and interpretations of possible differences in performance.

Some cultures promote independence of individual members, while others stress interdependence (McCollum and McBride, 1997). These cultural tendencies affect parents' socialization goals for their children and subsequent parent-child interactions. Parent-child interactions are among the most important daily activities for infants and toddlers, as they affect the organization of daily experiences and routines and contribute to cognitive, emotional, social and communication skills development in young children. In Russia, parents have more physical contacts with their children and are more protective than American parents (Press, 1989). These behaviours are consistent with the cultural goal of encouraging their young children to be dependent on adults, which is in contrast to the American culture's focus on increased autonomy.

Whether a culture is focused on survival or achievement also affects parenting. In cultures where health and subsistence are assured, parents tend to stress exploration and self-maintenance. In cultures where health and survival are not stable and dependable, parents tend to be more protective of their children's movements. This cultural differences may also contribute to variations in performance on measures such as the Bayley Scales. Culturally sensitive assessment involves consideration of the culture and region within which children live, as well as the goals the culture has for its children (Anderson, 1989). Practitioners using measures such as the Bayley Scales need to develop this cultural sensitivity and incorporate it into their work with young children and families. Cultural differences also exist in terms of views of child development (McCollum and McBride, 1997). Parents have different views about when their children should reach developmental milestones, such as walking or toileting. Parent-child interactions emphasize the skills

considered important for that time in that particular culture. Parents also hold different beliefs about how their children develop these skills. Some cultures believe that children develop naturally, without outside help or influence, while others view development as being shaped by society or as interactional in nature. These differences in beliefs contribute to differences in parenting styles and participation in child development activities or programs.

Differences in conceptualizations of health and disabilities are also found across cultures (Anderson, 1989). Hispanic and Native American cultures tend to associate health with religion, while many Asians believe that health problems and disabilities result from sins committed by relatives. These differences may contribute to differences in information provided during assessments, as well as variations in levels of understanding disabilities and well as their tendencies to seek and participate in treatment.

Another important area for psychologists to consider involves differences in national goals for early intervention. In Sweden, parents and professionals believe that a child-centered, creative environment is important for early childhood programs (Carlson and Stenmalm, 1989). Although professionals in the United States also support this view, parents in the United States value more formal academic tasks and conformity to rules as important goals for preschoolers. Similar differences have been found between the views of Finnish and American parents and educators, with Americans valuing a more formal and structured instructional approach (Hoot et al., 1996). In addition, more congruence has been found between parents and educators in Sweden and Finland than in the United States. The level of agreement may influence the philosophy of early intervention and resulting programs. These cultural variations contribute to differences in national goals for young children as well as differences in performance of varying skills and abilities, all of which may influence interpretations of results of developmental assessments. The use of tests such as the Bayley Scales may be most applicable in countries that focus on early intervention, which may not be the case with many underdeveloped, third world countries. Because of this, it may be preferable to establish local norms.

The significant discrepancies between test scores found in the present sample, while not surprising, provide important information for psychologists and other professionals involved in the assessment of infants and toddlers as they interpret scores on the new test. This information may be pertinent for those children whose performance is evaluated longitudinally with the Bayley Scales, such as those who are profoundly delayed, as scores on the old and new versions of the test are not interchangeable. In other words, when comparing results obtained with the BSID-II to scores obtained on previous testing with the BSID, it may appear that these children's skills have significantly declined.

In light of the validity and reliability research with the original Bayley Scales, it should be noted that the predictive validity of both versions of

the test is still questionable. According to Bayley, the abilities undergoing the most rapid development during early childhood appear to be most predictive of later cognitive functioning (Burns et al., 1992; Lewis et al., 1986). The predictive validity of the scales has been questioned in the literature (Burns et al., 1992) and Bayley has even questioned the ability of developmental testing to predict the relationship between ability in early childhood and later intelligence (Bayley, 1993). While the Bayley Scales do exist as valuable tools for the assessment of young children, they should be used with caution when attempting to predict future cognitive abilities.

The results of the present study should serve as the impetus for further research with the BSID-II. This study was limited in that it only included healthy, Caucasian children from intact families of middle to upper-middle socioeconomic status. It is recommended that the revised Bayley Scales be utilized in research studies with a variety of children from various ethnic and cultural backgrounds and clinical populations. Future studies should also involve the use of multiple examiners, as the use of a sole examiner may have made the present study subject to examiner effects. In the United States, legislation requiring the provision of preschool services has increased the demand for reliable measures of development in young children. The need for accurate assessment instruments for young children may not be required in other countries. It is expected that during the next several years, many studies of the revised Bayley Scales of Infant Development will be found in the literature both in America and across cultures attesting to its strength as a valuable measurement tool and that the revision will experience the same degree of popularity as the original measure.

REFERENCES

- Anderson, P. P. (1989) 'Issues in Serving Culturally Diverse Families of Young Children with Disabilities', *Early Child Development and Care* 50:167-88.
- Aylward, G. P., Verhulst, S.J., Bell, S. and Gyurke, J. S. (1995) 'Cognitive and Motor Score Differences in Biologically At-Risk Infants', *Infant Behavior and Development* 18: 43-52.
- Bayley, N. (1969) *Bayley Scales of Infant Development*. San Antonio, TX: Psychological Corporation.
- Bayley, N. (1993) *Bayley Scales of Infant Development, Second Edition*. San Antonio, TX: Psychological Corporation.
- Bracken, B. A. (1988) 'Ten Psychometric Reasons Why Similar Tests Produce Dissimilar Results', *Journal of School Psychology* 26: 155-166.
- Burns, W. J., Burns, K A. and Kabacoff, R. 1. (1992) 'Item and Factor Analyses of the Bayley Scales of Infant Development', *Advances in Infancy Research* 7: 199-214.
- Campbell, S. K, Siegel, E., Parr, C. A. and Ramey, C. T. (1986) 'Evidence for the Need to Reform the Bayley Scales of Infant Development Based on the Performance of a Population-Based Sample of 12-Month-Old Infants', *Topics in Early Childhood Special Education* 6: 83-96.
- Carlson, H. L. and Stenmalm, L. (1989) 'Professional and Parent Views of Early Childhood Programs: A Cross-Cultural Study', *Early Child Development and Care* 50: 51-64.
- Cook, M. J., Holder-Brown, L., Johnson, L. J. and Kilgo, J. L. (1989) 'An Examination of the Stability of the Bayley Scales of Infant Development with High-Risk Infants', *Journal of Early Intervention* 13: 45-9.
- Crowe, T. K, Deitz, J. C. and Bennett, F. C. (1987) 'The Relationship Between the Bayley Scales of Infant Development and Preschool Gross Motor and Cognitive Performance', *The American Journal of Occupational Therapy* 41: 374-78.
- Dilworth, A. and French, J. (1990) 'Development of a Cognitive Ability Test for 2-Year-Olds with Motor or Speech Delays', *Journal of Psychoeducational Assessment* 8: 42-50.
- Escalona, S. K (1984) , Social and Other Environmental Influences on the Cognitive and Personality Development of Low Birthweight Infants', *American Journal of Mental Deficiency* 88: 508-12.
- Flynn, J. R. (1984) 'The Mean IQ of Americans: Massive Gains 1932 to 1978', *Psychological Bulletin* 95: 29-51.

Flynn, J. R. (1987) 'Massive IQ Gains in 14 Nations: What IQ Tests Really Measure', *Psychological Bulletin* 95: 29-51.

Gerken, K C., Eliason, M. J. and Arthur, C. R. (1994) 'The Assessment of At-Risk Infants and Toddlers With the Bayley Mental Scale and the Battelle Developmental Inventory: Beyond the Data', *Psychology in the Schools* 31: 181-7.

Goldstein, D. J., Fogle, E. E., Wieber, J. L. and O'Shea, T. M. (1995) 'Comparison of the Bayley Scales of Infant Development-Second Edition and the Bayley Scales of Infant Development with Premature Infants', *Journal of Psychoeducational Assessment* 13: 391-6.

Greenfield, P.M. and Suzuki, L.K (1989) 'Culture and Human Development: Implications for Parenting, Education, Pediatrics and Mental Health', in I.C. Siegel and K A. Renninger (eds) *Handbook of Child Psychology, Volume 4* (pp. 1059-109), New York: Wiley.

Hoot, J. L., Parmar, R. and Hujala-Huttonen, E. (1996) 'A Comparison of The Views of Finnish and US Teachers, Administrators and Parents on Appropriate Educational Practice for Young Children', *Scandinavian Journal of Educational Research* 40: 349-66.

Jayashankarappa, B. S. (1986) 'Intellectual Tests and Social-Adaptive Behavioural Scales Used for the Assessment of the Mentally Handicapped in India', *Journal of Personality and Clinical Studies* 2: 89-98.

McCollum, J. A. and McBride, S. L. (1997) 'Ratings of Parent-Infant Interaction: Raising Questions of Cultural Validity', *Topics in Early Childhood Special Education* 17: 494-519.

Mellins, C. A., Levenson, Jr., R. L., Zawadzki, R., Kairam, R and Weston, M. (1994) 'Effects of pediatric HIV Infection and Prenatal Drug Exposure on Mental and Psychomotor Development', *Journal of Pediatric Psychology* 19: 617-28.

Nellis, L. and Gridley, B. E. (1994) 'Review of the Bayley Scales of Infant Development-Second Edition', *Journal of School Psychology*, 32: 201-9.

Press, B. K (1989) 'Observations on Family Life, Child Rearing and Education in the Soviet Union Today', *Early Child Development and Care* 50: 11-23.

Rubin, R. A. and Balow, B. (1979) 'Measures of Infant Development and Socioeconomic Status as Predictors of Later Intelligence and School Achievement', *Developmental Psychology* 15: 225-7.

Siegel, L. S., Cooper, D. C., Fitzhardinge, P. M. and Ash, A. J. (1995) 'The Use of the Mental Development Index of the Bayley Scale to Diagnose Language Delay in 2-Year-Old High-Risk Infants', *Infant Behavior and Development* 18: 483-6.

Snyder, S. and Sheehan, R. (1992) 'Rasch Analysis of the Standardization Data of the Bayley Mental Scale of Infant Development', *Diagnostique* 17: 185-94.

Tardif, T., Shatz, M. and Nailes, L. (1997) 'Caregiver Speech and Children's Use of Nouns versus Verbs: A Comparison of English, Italian and Mandarin', *Journal of Child Language* 24: 535-65.

Tasbihsazan, R., Nettelbeck, T. and Kirby, N. (1997) 'Increasing Mental Development Index in Australian Children: A Comparative Study of Two Versions of the Bayley Mental Scale', *Australian Psychologist* 32: 120-5.

Thompson, B., et al. (1994) 'The Validity of Mental and Motor Scores from the New Bayley Scales of Infant Development-II: A Second-Order Factor Analysis', paper presented at the Annual Meeting of the Southwest Educational Research Association. (ERIC Document Reproduction Service No. ED 366623)

van der Meulen, B. F. and Smrkovsky, M. (1985) 'Factor Analyses of Bayley's Infant Behavior Record: A Dutch Replication and Extension', *British Journal of Developmental Psychology* 3: 345-52.

Whaley, S. E., Sigman, M., Espinosa, M. P. and Neumann, C. G. (1998) 'Infant Predictors of Cognitive Development in an Undernourished Kenyan Population', *Developmental and Behavioral Pediatrics* 19: 1699-177.