

# Concurrent Validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales-2 in Children with Developmental Delays

Beth Provost, PhD, PT, Sandra Heimerl, MS, PT, Cate McClain, MD, PT, Nae-Hwa Kim, BA, Brian R. Lopez, PhD, and Piyadasa Kodituwakku, PhD

University of New Mexico, Health Sciences Center, Department of Orthopaedics Physical Therapy Program (B.P.) and Center for Development and Disability (S.H., C.M., N.K., B.L., P.K.), Albuquerque, New Mexico

**Purpose:** The purpose of this study was to explore concurrent validity of the age equivalent and standard scores of the Bayley Scales of Infant Development II (BSID II) Motor Scale and the Peabody Developmental Motor Scales-2 (PDMS-2), including correlations and clinical agreement between the scores of the two tests. **Methods:** One hundred ten children aged three to 41 months who were referred to an early childhood evaluation program because of concerns about their development were administered both the BSID II Motor Scale and the PDMS-2 as part of their developmental evaluations. **Results:** The correlation coefficients were high to very high for age-equivalent scores, and the Locomotion Subscale had the closest agreement with the BSID II Motor Scale age equivalent. The correlation coefficients were moderate to high for standard scores, and there was only slight agreement between the tests for standard score categories. More than 75% of the 70 children in this study whose scores on the BSID II supported eligibility for services based on scores at least two SD below the mean of the test would not have qualified for services if the PDMS-2 standard scores alone were used to assess their eligibility. Approximately half the children who showed appropriate total motor performance on the PDMS-2 were classified as delayed on the BSID II Motor Scale. **Conclusions:** The study supports concurrent validity of the tests only for certain subscale age-equivalent scores, particularly the BSID II Motor Scale with the PDMS-2 Locomotion Subscale. The current findings suggest that the standard scores show poor agreement and have low concurrent validity. There are marked differences in the standard scores of the two tests that may affect a child's eligibility for services in some states, and therapists should be cautious when making clinical decisions based solely on standard scores of one test. (*Pediatr Phys Ther* 2004;16:149–156) **Key words:** child development, child, comparative study, developmental disabilities/diagnosis, physical therapy/methods, motor skills, neuropsychological tests/standards, psychometrics

## INTRODUCTION

Physical therapists are frequently called on to perform evaluations of young children with suspected or known

risk of developmental delay, typically in gross and fine motor development.<sup>1</sup> These evaluations may serve several important purposes for children: to identify delays or disorders in development, to plan intervention activities, and to monitor children's progress. The evaluations may also provide data on which decisions will be made regarding children's eligibility for certain early intervention services.<sup>1</sup> Although the criteria for eligibility vary among states, most criteria are based on scores from standardized tests that document delays in development, calculated from either age-equivalent scores or standard scores. Of standardized tests, the Motor Scale of the Bayley Scales of Infant

0898-5669/04/1603-0149  
Pediatric Physical Therapy  
Copyright © 2004 Lippincott Williams & Wilkins, Inc.

Address correspondence to: Beth Provost, PhD, PT, Health Sciences Center, Physical Therapy Program, MSC09 5230, 1 University of New Mexico, Albuquerque, NM 87131-0001. Email: eprovost@salud.unm.edu

DOI: 10.1097/01.PEP.0000136005.41585.FE

Development II (BSID II)<sup>2</sup> and the Peabody Developmental Motor Scales-2 (PDMS-2)<sup>3</sup> are two of the most commonly used discriminative measures in early intervention.<sup>1</sup>

Discriminative measures distinguish between children who have and do not have a particular characteristic.<sup>4</sup> Physical therapists must be able to distinguish between children who are developing typically in their motor performance and children who are developing atypically to reassure the parents of children who are typically developing and to monitor the development of children whose performance is different from that of their peers. Physical therapists must also be able to distinguish between children who have significant motor delays and those with little or no delays to know which children may be eligible for intervention services in their community. Physical therapists use their clinical judgment to make decisions about treatment needs, and their judgment is often assisted by results of standardized testing. If children's age-equivalent scores are close to their chronological age (or age adjusted for prematurity if appropriate) and if their standard scores are appropriately close to the mean of the test, then therapists can be confident that children are developing appropriately for their age. If children's age-equivalent scores are below their chronological age and if their standard scores are at least two SD below the mean of the test, then therapists can be confident that the children are not developing appropriately for their age and may benefit from services.

Physical therapists must choose tests that yield valid measurements of children's motor performance for both age-equivalent scores and standard scores. Both the BSID II Motor Scale and the PDMS-2 have appropriate psychometric properties of reliability and validity, are standardized and norm referenced, and measure motor skills in young children.<sup>2,3</sup> Differences, however, exist between the tests that may influence a therapist's choice of one test over the other. Benefits of choosing the BSID II include compatibility with the BSID II Mental Scale, little need for equipment, and a combined score for both fine and gross motor areas. Benefits of choosing the PDMS-2 include separate scores for fine and gross motor skills and more test items in the fine motor area. However, if a young child's motor performance is evaluated with one or the other of these tests, an important clinical question is raised: Are the child's scores on these two standardized tests similar and do they lead to the same conclusions about his or her development being typical or delayed?

Concurrent validity refers to the degree to which scores on a test are related to scores on another test administered at the same time.<sup>5</sup> Concurrent validity has been reported for the original BSID Motor Scale and the original PDMS,<sup>6</sup> as well as for the BSID II Motor Scale and the original PDMS.<sup>7</sup> Provost et al<sup>7</sup> found evidence for the concurrent validity of the BSID II Motor Scale and the PDMS for age-equivalent scores but not for standard scores in two-year-old Native American children. However, the authors stated that further studies of the same relationship are needed with a larger sample from diverse ethnic back-

grounds as well as with children who are at risk of or who have developmental delays. No concurrent validity studies have been published for the BSID II Motor Scale and the PDMS-2.

The purpose of this study was to explore correlations and clinical agreement between the age-equivalent and standard scores of the BSID II Motor Scale and the PDMS-2 to address the following questions:

1. Do age-equivalent scores on the BSID II Motor Scale correlate with age-equivalent scores on the PDMS-2 Gross Motor Subscales and the PDMS-2 Fine Motor Subscales?
2. Do standard scores on the BSID II Motor Scale [Psychomotor Development Index (PDI)] correlate with standard scores on the PDMS-2 (Gross Motor Quotient, Fine Motor Quotient, and Total Motor Quotient)?
3. Do age-equivalent scores and standard scores on the BSID II Motor Scale and the PDMS-2 agree for clinically important criteria?
4. Do the BSID II Motor Scale and the PDMS-2 categorize children similarly for typical motor performance as well as for significant delays?

## METHODS

### Participants

One hundred ten children (age: mean = 25.3 months, SD = 9.7, range = 3–41) were recruited from the University of New Mexico Health Sciences Center's Center for Development and Disability Early Childhood Evaluation Program (ECEP) to participate in this study. Infants and children throughout New Mexico are referred to the Center for Development and Disability ECEP for comprehensive interdisciplinary developmental evaluations. This well-established university program (University Center for Excellence in Developmental Disabilities, Education, Research, and Service) has been serving young children and their families for more than 20 years, and the program evaluates approximately 300 children per year between the ages of two and 48 months. Children are referred to the program because they are at risk of delay or have an established developmental delay and often have complex neurodevelopmental problems. The interdisciplinary team that sees each child is composed of a cognitive specialist who assesses cognitive, adaptive, and social/emotional development; a speech language pathologist who assesses speech and language development; a physical therapist or an occupational therapist who assesses gross and fine motor development and sensory processing abilities; and a pediatrician who conducts a medical examination. Standardized testing is routinely done to identify developmental levels and assess the quality of the child's skills in the various areas of development. The motor assessment routinely includes the administration of the BSID II Motor Scale and/or the PDMS-2.

This study was approved by the University of New Mexico School of Medicine's Human Research Review

Committee. Each child's parent or legal guardian was informed about the study and signed the consent form on the day of the clinic visit. No compensation was provided for participants' involvement in this study.

The participants were 75 boys and 35 girls. The ethnicity of the sample was 35.5% white, 52.7% Hispanic, 9.1% Native American, 1.8% Asian, and 0.9% African-American/Hispanic. Eighty-seven of the children (79%) had developmental delays in at least two areas of development (cognitive, speech, or motor delay), whereas 10 children (9%) had isolated speech delay and eight children (7%) had isolated motor delay. In addition to delays, 20 of the children (18%) had a diagnosis of autism spectrum disorder/pervasive developmental disorder.

## Procedures

One of two physical therapists (B.P., S.H.) administered the BSID II Motor Scale and the PDMS-2 to each participating child as part of their routine ECEP evaluation. During the routine evaluation, the physical therapist explained the study to the parent or legal guardian and answered any questions. If agreeable, the primary caregiver signed the consent form to have the child's scores on the two motor tests entered into the database. If parents elected not to participate, the child continued to receive the routine comprehensive evaluation. The two experienced pediatric physical therapists involved in this study have performed developmental testing of young children for more than 25 years each. Interrater reliability was established between the two therapists on six children. Three children were videotaped and then scored from the videotape by both researchers. Three children were tested and scored in the clinic setting by one researcher while being observed and scored at the same time by the other researcher. Interrater agreement on the six children averaged 91% for the PDMS-2 (range, 85–98%), and 96% for the BSID II Motor Scale (range, 88–100%).

**BSID.** The BSID was revised to become the BSID II in 1993, and normal values were redetermined on a sample of 1700 children between birth and 42 months.<sup>2</sup> The BSID II consists of a Mental Scale, a Motor Scale, and a Behavior Rating Scale and has been reviewed for psychometric strengths and limitations.<sup>8</sup> According to the manual, the Motor Scale assesses degree of control of the body, coordination of the large muscles, finer manipulatory skills of the hands and fingers, dynamic movement, dynamic praxis, postural imitation, and stereognosis. Raw scores on the Motor Scale are converted to developmental age-equivalent scores as well as standard scores called the PDI scores. The mean standard score is 100 and the standard deviation (SD) is 15. The BSID II Motor Scale classifies performance based on whole-number SD from the mean, into categories of accelerated, within normal limits (WNL), mildly delayed, and significantly delayed.

**PDMS.** The PDMS was revised to become the PDMS-2 in 2000, and normal values were redetermined on a sample of 2003 children.<sup>3</sup> The PDMS-2 consists of six subtests: Reflexes (for children from birth through 11

months), Stationary (ability to sustain control of body within its center of gravity), Locomotion (ability to move from one place to another), Object Manipulation (ability to manipulate balls for children 12 months and older), Grasping (ability to use hands), and Visual-Motor Integration (ability to use visual perceptual skills to perform complex eye-hand coordination tasks). Raw scores on the PDMS-2 are converted to age-equivalent scores for the subtests, percentiles, subtest standard scores, and composite standard scores called motor quotients. The Reflexes or Object Manipulation, Stationary, and Locomotion subtests contribute to the Gross Motor Quotient, and the Grasping and Visual-Motor Integration subtests contribute to the Fine Motor Quotient, and the Total Motor Quotient is formed by a combination of the results of the gross and fine motor subtests. Although the PDMS-2 has a mean motor quotient standard score of 100 and SD of 15, it classifies performance primarily based on 10-point increments (rather than the 15-point SD increments) into categories of very superior, superior, above average, average, below average, poor, and very poor. Table 1 compares the standard score classifications of both tests.

## Data Analysis

Each child's age-equivalent scores and the standard scores were calculated for the BSID II Motor Scale and the PDMS-2. The scores were entered into a research database (Microsoft Excel) and analyzed using SPSS 11.5 for Windows. Concurrent validity was examined using correlational analysis, frequency of agreement, and statistical tests of symmetry and kappa statistics.

Pearson product moment correlation coefficients were calculated between the scores of the two tests. Correlation coefficients are used to quantitatively describe the strength and direction of a relationship between two variables, and the Pearson product-moment coefficient is the most commonly reported measure of correlation, based on the concept of covariance.<sup>9</sup> The probability of the correlation occurring by chance is determined by statistical tests

**TABLE 1**

Comparison of Classifications of Standard Scores for the BSID II Motor Scale (Psychomotor Development Index Scores) and the PDMS-2 (Motor Quotients)

BSID II PDI Scores		PDMS-2 Motor Quotients	
Standard Scores	Classifications	Standard Scores	Classifications
		131–165	Very superior
		121–130	Superior
≥115	Accelerated	111–120	Above average
85–114	Within normal limits	90–110	Average
70–84	Mildly delayed	80–89	Below average
		70–79	Poor
≤69	Significantly delayed	35–69	Very poor

BSID II = Bayley Scales of Infant Development II; PDMS-2 = Peabody Developmental Motor Scales-2.

and is denoted as  $p$ ; therefore,  $p$  values also were calculated.<sup>9</sup> The strength of the correlation coefficient was categorized according to Munro:<sup>10</sup> 0 to 0.25 = little if any correlation, 0.26 to 0.49 = low correlation, 0.50 to 0.69 = moderate correlation, 0.70 to 0.89 = high correlation, and 0.90 to 1.0 = very high correlation.

The degree of similarity of the scores on each test, for both age equivalents and standard scores, is an important clinical issue for pediatric physical therapists. However, correlation provides information only relative to the order of the scores, and it does not provide any information relative to the magnitude of the difference between sets of data (eg, scores on two tests).<sup>10</sup> Therefore, frequency of agreement measures were also used to help determine how comparable the actual scores were on both tests, and these measures included percentages of agreement and statistical tests for symmetry and kappa statistics. Frequency of agreement between age-equivalent scores in months was calculated between the BSID II Motor Scale and the PDMS-2 Subscales as percentages of exact agreement and percentages of PDMS-2 scores that were within  $\pm 1$ ,  $\pm 2$ ,  $\pm 3$ ,  $\pm 4$ ,  $\pm 5$ , and  $\pm 6$  months from the BSID II age-equivalent scores. For example, if a child had an age-equivalent score of 20 months on the BSID II Motor Scale, and an age-equivalent score of 23 or 17 months on the PDMS-2 Locomotion Subscale, his or her PDMS-2 score would be calculated as within  $\pm 3$  months of the BSID II age equivalent.

The agreement of the standard scores was calculated for the categories used by both tests. Because only one child scored higher than average or WNL, the categories WNL and average were collapsed with categories above that classification for each test. A clinically important question for therapists is whether children are delayed enough to merit special services. Both tests have classifications at least two SD below the mean of the test (standard score  $\leq 69$ ), called significantly delayed on the BSID II and very poor on the PDMS-2. Frequency of agreement was calculated on how often there was agreement between the tests for a score at least two SD below the mean of the tests because that is a criterion for eligibility for intervention services in some states. Another clinically important question is whether children perform typically compared with other children their age. Frequency of agreement was calculated on how often there was agreement between a score of WNL or above (a standard score of  $\geq 85$ ) on the BSID II Motor Scale and a score of average or above (a standard score of  $\geq 90$ ) on the PDMS-2.

## RESULTS

The correlation and frequency of agreement results for the age-equivalent scores are presented first, followed by the standard score results.

### Age-Equivalent Scores

**Correlation findings.** Pearson product-moment correlations were calculated to determine the concurrent validity between the BSID II Motor Scale and the PDMS-2

Subscales for age-equivalent scores. Table 2 presents the correlations, sample size, means, and SD for the age-equivalent scores. The Reflex Subscale is used with children younger than 12 months, and 13 children in the sample were younger than 12 months. The Object Manipulation Subscale is used with children 12 months and older, and two older children were unable to receive an age-equivalent score in Object Manipulation because they had a raw score of 0 on the subscale. According to Munro's standards,<sup>10</sup> the coefficients were high to very high for all correlations.

**Frequency of agreement.** Frequency of agreement between age-equivalent scores is presented in Table 3. The Reflexes and the Locomotion Subscales had 96% to 100% agreement within three months. The Stationary, Object Manipulation, and Visual-Motor Integration Subscales showed 90% to 95% agreement within five months, and the Grasping Subscale reached 90% agreement within six months.

### Standard Scores

**Correlation findings.** Pearson product-moment correlations were calculated to determine the concurrent validity between the standard scores of the BSID II and the PDMS-2. The BSID II PDIs were correlated with the PDMS-2 Gross Motor Quotients, Fine Motor Quotients, and Total Motor Quotients. Table 4 presents the correlations, means, and SD for the standard scores. There were 110 standard scores for each correlation. According to Munro's<sup>10</sup> standards, the coefficients were moderate to high for all correlations.

**Frequency of agreement.** Tables 5 through 7 present the numbers of children whose standard scores agreed in the various categories used by both tests. Because only one child (with a score of above average on the PDMS-2 Gross Motor Quotient) scored higher than average or WNL, the categories WNL and average were collapsed with categories above those classifications. The bolded numbers on the diagonal are in the same general grouping on both tests.

**TABLE 2**

Pearson Product Moment Correlations for the Age-Equivalent Scores on the BSID II Motor Scale and the PDMS-2 Subscales, and Number of Children, Mean Age-Equivalent Scores in Months, and Standard Deviations for Each Subscale

	<i>r</i> with BSID II	N	Mean	SD
BSID II Motor Scale	—	110	18.9	8.5
PDMS-2 subscales				
Reflexes	0.89	13	5.0	2.5
Stationary	0.93	110	17.6	7.6
Locomotion	0.97	110	18.9	8.6
Object manipulation	0.90	95	21.6	6.3
Grasping	0.85	110	18.8	8.8
Visual/motor Integration	0.94	110	20.0	8.8

Note: All the correlations were highly significant,  $p < 0.001$ . BSID II = Bayley Scales of Infant Development II; PDMS-2 = Peabody Developmental Motor Scales-2.

**TABLE 3**

Percentages of Agreement (Exact Agreement and Agreement Within One Month Through Agreement Within Six Months) Between the Age-Equivalent Scores of the BSID II Motor Scale and the PDMS-2 Subscales

	Exact	±1 mo	±2 mo	±3 mo	±4 mo	±5 mo	±6 mo
Reflexes	38	77	<b>100</b>				
Stationary	12	34	55	69	79	<b>90</b>	<b>96</b>
Locomotion	23	55	79	<b>96</b>	<b>97</b>	<b>99</b>	<b>100</b>
Object manipulation	16	31	56	75	84	<b>95</b>	<b>97</b>
Grasping	13	29	46	61	74	<b>81</b>	<b>90</b>
Visual/motor integration	15	39	59	78	88	<b>93</b>	<b>95</b>

Note: Bolded indicated numbers ≥90% agreement between scores. BSID II = Bayley Scales of Infant Development II; PDMS-2 = Peabody Developmental Motor Scales-2.

**TABLE 4**

Pearson Product Moment Correlations for the Standard Scores on the BSID II Motor Scale (PDI Scores) and the PDMS-2 (Motor Quotients), and Mean Standard Scores and Standard Deviations

	<i>r</i> with BSID II	Mean	SD
BSID II Motor Scale (PDI)	—	65.6	16.1
PDMS-2			
Gross Motor Quotient	0.75	82.8	11.0
Fine Motor Quotient	0.67	87.0	11.9
Total Motor Quotient	0.76	83.0	11.5

Note: All the correlations were significant,  $p < 0.001$ . BSID II = Bayley Scales of Infant Development II; PDI = Psychomotor Development Index; PDMS-2 = Peabody Developmental Motor Scales-2.

**Classifications of significantly delayed versus very poor.** All children who scored very poor on the PDMS-2 also scored significantly delayed on the BSID II. However, of the 70 children whose standard scores were in the significantly delayed category on the BSID II Motor Scale, only 14 children scored very poor on the PDMS-2 Gross Motor Quotient, only seven children scored very poor on the PDMS-2 Fine Motor Quotient, and only 16 children scored very poor on the PDMS-2 Total Motor Quotient. Therefore, only 10% to 23% of the sample scoring at least two SD below the mean of the BSID II Motor Scale scored comparably on the PDMS-2. More than 75% of the children who were significantly delayed on the BSID II Motor Scale were not very poor on the PDMS-2 and were thus categorized differently for these clinically important criteria.

**Classifications of WNL or above versus average or above.** Of the 20 children whose standard scores were in the WNL or above category on the BSID II Motor Scale, 16 children scored average or above on the PDMS-2 Gross Motor Quotient, 19 children scored average or above on the PDMS-2 Fine Motor Quotient, and 18 children scored average or above on the PDMS-2 Total Motor Quotient. Therefore, for those children who were categorized as developing appropriately in their motor skills on the BSID II Motor Scale, 80% to 95% were also categorized as at least typically developing on the PDMS-2.

Of the 27 children who scored average or above on the PDMS-2 Gross Motor Quotient, 16 scored WNL or above on the BSID II Motor Scale. Of the 46 children who scored average or above on the PDMS-2 Fine Motor Quotient, 19

scored WNL or above on the BSID II Motor Scale. Of the 34 children who scored average or above on the PDMS-2 Total Motor Quotient, 18 scored WNL or above on the BSID II Motor Scale. Therefore, for children who were categorized as developing appropriately in their motor skills on the PDMS-2, only 41% to 59% were categorized comparably on the BSID II Motor Scale, and approximately half the children who were typical on the PDMS-2 were classified as delayed on the BSID II Motor Scale.

**Kappa statistics and McNemar’s test.** The degree of agreement between the BSID-II Motor Scale and the PDMS-2 standard scores was also tested by computing the kappa coefficient,<sup>11</sup> which is a measure of association corrected for chance agreement. The simple kappas for the BSID II Motor Scale and the PDMS-2 Gross Motor, Fine Motor, and Total Motor Quotient scores were 0.13, 0.02, and 0.09, respectively. These kappa values indicate only slight agreement between the two tests.<sup>12</sup> Tests of symmetry (McNemar’s test) showed significant differences in classification on PDMS-Gross Motor [ $\chi^2(1) = 54.01, p = 0.0001$ ], Fine Motor [ $\chi^2(1) = 61.02, p = 0.0001$ ], and Total Motor [ $\chi^2(1) = 52.01, p = 0.0001$ ] scores. As Table 5 illustrates, 35 children classified as below average or average or above in gross motor behavior on the PDMS-2 were assessed as significantly delayed on the BSID II Motor Scale. Forty-three children rated to be below average or average or above in fine motor behavior on the PDMS-2 were classified as significantly delayed on the BSID II Motor Scale (Table 6). Furthermore, the BSID II Motor Scale classified 34 children with below average to average or above Total Motor Quotient scores on the PDMS-2 as significantly delayed (Table 7).

**DISCUSSION**

The results of this study raise concerns about the concurrent validity of the BSID II Motor Scale and the PDMS-2 for clinically important issues and reveal a complex relationship between the scores of the two tests. Pediatric physical therapists who are using these tests to help them make decisions about a child’s need for services based on age-equivalent scores or standard scores must be aware of the potential differences in outcomes for some children evaluated with these tests. Depending on the test used, a child could be determined eligible for or not eligible for early intervention services in their community.

**TABLE 5**

Number of Children with PDMS-2 Gross Motor Quotients and BSID II Motor Scale PDI Scores Grouped into Categories Used by Both Tests

		PDMS-2 Gross Motor Quotients				
		Average or Above ≥90	Below Average		Poor 70-79	Very Poor ≤69
			85-89	80-84		
BSID II PDI						
WNL or above	≥90	<b>9</b>	0	0	0	0
WNL	85-89	7	4	0	0	0
Mild delay	80-84	3	3	1	0	0
Mild delay	70-79	6	3	1	2	0
Significant delay	≤69	2	19	14	21	14

Note: Bolded numbers on the diagonal are in the same standard score groups on both tests. PDMS-2 = Peabody Developmental Motor Scales-2; BSID II = Bayley Scales of Infant Development II; PDI = Psychomotor Development Index; WNL = within normal limits.

**TABLE 6**

Number of Children with PDMS-2 Fine Motor Quotients and BSID II Motor Scale PDI Scores Grouped into Categories Used by Both Tests

		PDMS-2 Fine Motor Quotients				
		Average or Above ≥90	Below Average		Poor 70-79	Very Poor ≤69
			85-89	80-84		
BSID II PDI						
WNL or above	≥90	<b>8</b>	1	0	0	0
WNL	85-89	11	0	0	0	0
Mild delay	80-84	7	1	0	0	0
Mild delay	70-79	8	2	1	1	0
Significant delay	≤69	12	24	7	20	7

Note: Bolded numbers on the diagonal are in the same standard score groups on both tests. PDMS-2 = Peabody Developmental Motor Scales-2; BSID II = Bayley Scales of Infant Development II; PDI = Psychomotor Development Index; WNL = within normal limits.

**TABLE 7**

Number of Children with PDMS-2 Total Motor Quotients and BSID II Motor Scale PDI Scores Grouped into Categories Used by Both Tests

		PDMS-2 Total Motor Quotients				
		Average or Above ≥90	Below Average		Poor 70-79	Very Poor ≤69
			85-89	80-84		
BSID II PDI						
WNL or above	≥90	<b>8</b>	1	0	0	0
WNL	85-89	10	1	0	0	0
Mild delay	80-84	3	5	0	0	0
Mild delay	70-79	7	1	2	2	0
Significant delay	≤69	6	14	14	20	16

Note: Bolded numbers on the diagonal are in the same standard score groups on both tests. PDMS-2 = Peabody Developmental Motor Scales-2; BSID II = Bayley Scales of Infant Development II; PDI = Psychomotor Development Index; WNL = within normal limits.

When tests are used in important decision making, correlations between two forms of the same test must be very high, approximately 0.95.<sup>10</sup> The same reasoning could be applied to use of the BSID II Motor Scale and the PDMS-2, where important clinical decisions, such as eligibility for services or need for monitoring development, depend on the results. It was encouraging that the age-equivalent scores on BSID II Motor Scale showed very high correlations with age-equivalent scores on four of six subscales of the PDMS-2. In particular, the Locomotion Subscale of the PDMS-2, which measures gait and movement abilities in young children and therefore is very useful for physical therapists, showed an impressive  $r = 0.97$  correlation with the BSID II Motor Scale. In addition to most of

the correlations between the age-equivalent scores on both tests being very high, the age-equivalent score means of the PDMS-2 subscales (17.6-21.6 months) were relatively close to the BSID II Motor Scale mean of 18.9 months for the 110 participants, and the Locomotion Subscale mean was the same as the BSID II mean.

However, it is of some concern that the actual agreement between the age-equivalent scores on the BSID II Motor Scale and the PDMS-2 was lower than expected, and these differences have the potential to affect decisions about a child's need for services. In addition to the highest correlation and the closest mean, the Locomotion Subscale showed most agreement in actual age-equivalent scores with the BSID II Motor Scale (96% agreement within three

months). However, the Stationary, Object Manipulation, and Visual-Motor Integration Subscales showed 90% to 95% agreement only within five months of the BSID II Motor Scale age equivalent, and the Grasping Subscale reached 90% agreement only within six months. The implications of these differences are exemplified in a 20-month-old child who has a 15-month age equivalent on the BSID II Motor Scale. This represents a 25% delay and meets eligibility criteria for intervention services in some states. A difference of five months in a PDMS-2 subscale age equivalent (eg, 20-month age equivalent) might indicate 0% delay and therefore would not meet eligibility criteria.

In contrast to age-equivalent scores, the correlations between the standard scores for the BSID II Motor Scale and PDMS-2, although statistically significant, were only moderate to high, and were substantially lower than the age-equivalent score correlations. None of the correlations were very high by Munro's<sup>10</sup> criteria, the standard scores had only slight agreement on the statistical measures, and clinical differences between the outcomes of the two tests were of concern. The mean of the BSID II Motor Scale standard scores for the 110 children in this study was 65.6 (at least two SD below the mean of the test) compared with the PDMS-2 motor quotient standard scores of 82.8 to 87.0 (from slightly less than one SD to slightly more than one SD below the mean of the test), despite the fact that both tests have the same mean of 100 and SD of 15.

A clinically important question for therapists is whether a child is delayed enough to merit special services in their community, such as having a standard score at least two SD below the mean of the test, a common criterion for eligibility. Physical therapists might assume that a child who was classified as significantly delayed on the BSID II Motor Scale would also be classified as very poor on the PDMS-2, but that assumption was not supported in this study. Although all children who scored very poor on the PDMS-2 also scored significantly delayed on the BSID II Motor Scale, the converse was not true. In fact, less than 25% of the sample scoring at least two SD below the mean on the BSID II Motor Scale also scored comparably on the PDMS-2. That discrepancy in evaluation outcomes means that more than 75% of the 70 children in this study whose scores on the BSID II supported eligibility for services based on scores at least two SD below the mean of the test would not have qualified for services if the PDMS-2 standard scores alone were used to assess their eligibility. This is problematic for therapists who are using the tests to help determine whether a child is eligible for services based on a significant delay of at least two SD below the mean in motor skills.

Another clinically important question for therapists is whether a child is performing typically compared with other children his age. Physical therapists might assume that a child who was classified as developing appropriately on the PDMS-2 would also be classified as developing appropriately on the BSID II Motor Scale, but that assumption was not supported in this study. Although 80% to 95% of the children whose standard scores were WNL or above on

the BSID II Motor Scale also scored average or above on the PDMS-2, the converse was not true. Only 41% to 59% of children who were categorized as developing appropriately in their motor skills on the PDMS-2 were categorized comparably on the BSID II Motor Scale. In fact, approximately half the children who showed appropriate total motor performance on the PDMS-2 were classified as delayed on the BSID II Motor Scale. This is problematic for therapists who may be using the tests to help determine whether a child's performance needs to be monitored, perhaps because of a medical history that puts them at risk of developmental delay. This study showed that a therapist's ability to reassure families of their children's appropriate motor development could be dependent on which test was administered to the children. In addition, it was disconcerting to realize that a wide discrepancy occurred for the children (two children for Gross Motor Quotient, 12 for Fine Motor Quotient, and six for Total Motor Quotient) who were categorized as average or above on the PDMS-2 but scored significantly delayed on the BSID II Motor Scale. For these children, the results of one test would suggest the need for intervention services, whereas the other test would suggest no service needs whatsoever.

Sensitivity is the ability of a test to identify correctly those who actually have a disorder, such as a motor delay, and specificity refers to the ability of a test to identify correctly those who do not have a disorder.<sup>13</sup> It is possible that the BSID II Motor Scale, which identified 70 of the 110 children in the study as significantly delayed, may have a higher sensitivity than the PDMS-2 (the PDMS-2 Total Motor Quotient categorized only 16 of the 110 children as very poor) for correctly identifying children with a delay. Conversely, the BSID II Motor Scale may be overidentifying children as delayed. It is possible that the PDMS-2 (which identified 34 the 110 children as average or above on the Total Motor Quotient) may have a higher specificity than the BSID II Motor Scale (which identified only 18 of the 110 children as WNL or above) for correctly identifying children who do not have a delay. Conversely, the PDMS-2 may be underidentifying delayed children. Whatever the case, the clinical results when using these two tests may be confusing to therapists unaware of these issues, and further research is needed to determine which classification truly represents the "correct" picture of the child.

Physical therapists administer motor tests to young children to assist them in their judgment of a child's motor abilities and need for services. The therapists may depend on a child's motor age-equivalent score to help them assess how many months behind the child is compared with his or her chronological age. The therapists may alternately depend on the child's standard score to help them assess whether his or her scores fall outside the realm of certain SD from the mean of the test, which would classify his or her score as delayed. The BSID II and the PDMS-2 are both useful tools for physical therapists when assessing young children. However, it is important for therapists to understand the strengths and limitations of any assessment tools that they choose to use with children. The results of this

study show that there are differences in the scores of the BSID II Motor Scale and the PDMS-2 that may affect clinically important decisions. Recommendations by the authors are for therapists to use these tests with awareness of the concurrent validity concerns. It is important for physical therapists to use their clinical judgment when assessing young children and to realize that test scores are useful in assisting their judgment but are not the only consideration. A therapist should use a test but not “be used” by the test results. If a therapist believes that a child would benefit from services, then the therapist should be aware that the child’s scores on the BSID II Motor Scale would more likely meet eligibility criteria for services than his or her scores on the PDMS-2. If a therapist believes that a child is developing appropriately, then the therapist should be aware that the child’s scores on the PDMS-2 would more likely support that clinical judgment.

## CONCLUSIONS

The BSID II Motor Scale and the PDMS-2 measure motor skills and identify motor delays in young children, and both tests have numerous strengths and pertinent applications for physical therapists. The results of this study provide clinically important information for physical therapists that are expected to accurately evaluate young children and determine whether a child is eligible for early intervention services because of delays in motor development or whether a child is developing appropriately. This study supports concurrent validity of the tests only for certain subscale age-equivalent scores, particularly the BSID II Motor Scale with the PDMS-2 Locomotion Subscale. The current findings suggest that the standard scores of the BSID II Motor Scale and the PDMS-2 show poor agreement and have low concurrent validity. There are

marked differences in the standard scores of the two tests that may affect a child’s eligibility for services in some states, and therapists should be guarded when making clinical decisions based solely on standard scores of one test.

## REFERENCES

1. Quinn L, Gordon J. *Functional Outcomes Documentation for Rehabilitation*. St. Louis: Elsevier Science; 2003.
2. Bayley N. *Bayley Scales of Infant Development Manual*, 2nd ed. San Antonio, TX: The Psychological Corporation; 1993.
3. Folio MR, Fewell RR. *Peabody Developmental Motor Scales Examiner’s Manual*, 2nd ed. Austin, TX: Pro-Ed; 2000.
4. McEwen I. *Providing Physical Therapy Services Under Parts B & C of the Individuals with Disabilities Education Act (IDEA)*. Alexandria, VA: American Physical Therapy Association; 2000.
5. Gay LR. *Educational Research*. Upper Saddle River, NJ: Prentice-Hall; 1996.
6. Palisano R. Concurrent and predictive validities of the Bayley Motor scale and the Peabody Developmental Motor Scales. *Phys Ther*. 1986; 66:1714–1719.
7. Provost B, Crowe TK, McClain C. Concurrent validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales in two-year-old children. *Phys Occup Ther Pediatr*. 2000;20:5–18.
8. Koseck K. Review and evaluation of the psychometric properties of the revised Bayley Scales of Infant Development. *Pediatr Phys Ther*. 1999;11:198–204.
9. Portney LG, Watkins MP. *Foundations of Clinical Research Applications to Practice*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall; 2000.
10. Munro BH. *Statistical Methods for Health Care Research*. Philadelphia: Lippincott-Raven; 1997.
11. Cohen J. A coefficient for agreement for nominal scales. *Educ Psychol Measure*. 1960;20:37–46.
12. Everitt BS. *The Analysis of Contingency Tables*, 2nd ed. London: Chapman & Hill; 1994.
13. Brenneman S. Tests of infant and child development. In: Tecklin J, ed. *Pediatric Physical Therapy*. Philadelphia: JB Lippincott; 1994: 24–55.