informa
healthcare

Concurrent Validity and Reliability of the Alberta Infant Motor Scale in Infants at Dual Risk for Motor Delays

Patricia Snyder, PhD
Jane M. Eason, PT, PhD
Darbi Philibert, MHS, LOTR
Andrea Ridgway, PhD
Tiffany McCaughey, MHS

ABSTRACT. Concurrent validity of scores for the Alberta Infant Motor Scale (AIMS) and the Peabody Developmental Gross Motor Scale-2 (PDGMS-2) was examined with a sample of 35 infants at dual risk for motor delays or disabilities. Dual risk was defined as low birthweight

At the time the study was conducted, all authors were affiliated with the Early Intervention Institute at Louisiana State University Health Sciences Center, New Orleans, LA. Patricia A. Snyder, PhD, is now David Lawrence Jr. Endowed Chair in Early Childhood Studies and Professor of Special Education, University of Florida, College of Education. Jane M. Eason, PT, PhD, is Associate Professor, Department of Physical Therapy, Louisiana State University Health Sciences Center– New Orleans; Darbi Philibert, MHS, LOTR, is now in private practice in New Orleans, LA; Andrea Ridgway, PhD, is now with Autism Spectrum Therapies in Arcadia, CA, Tiffany McCaughey, MHS, is now at the University of Illinois at Urbana-Champaign.

We thank the students from the pediatric research groups in the Master of Physical Therapy program at Louisiana State University Health Sciences Center, who assisted with data collection and the families and children who participated in this research. We particularly appreciate the support of Gina Kunz, PhD, who at the time of this study was the Director of the Early Childhood Development Clinic, Louisiana State University Health Sciences Center in New Orleans, LA.

Address correspondence to: Patricia A. Snyder, University of Florida, College of Education, G315 Norman Hall, PO Box 117050, Gainesville, FL 32611-7050 (E-mail: patriciasnyder@coe.ufl.edu).

Physical & Occupational Therapy in Pediatrics, Vol. 28(3), 2008
Available online at http://potp.haworthpress.com

doi: 10.1080/01942630802224892

($\leq$ 1,500 g at birth) and environmental risk (children from low-income families). Measures were administered to participants during one visit to a high-risk, follow-up program. Pearson product-moment correlation coefficients were high ($r = .90$ to .97), with the highest correlation between PDGMS-2 locomotion subscale and AIMS total score. Correlation coefficients were lower for infants > 9 months of age. Novice examiners' scores on both measures closely approximated those of experienced examiners (ICC range = .98 to .99). The results support concurrent validity of the AIMS and PDGMS-2 for infants at dual risk and have implications for using the AIMS in high-risk follow-up programs, particularly in relation to evaluating functional components of motor performance and ease of administration.

KEYWORDS. Alberta Infant Motor Scale, Peabody Developmental Gross Motor Scale, concurrent validity, inter-rater reliability, infants at dual risk

Early intervention practitioners, particularly pediatric physical and occupational therapists, frequently are involved in evaluations of the motor status of infants and toddlers suspected to be at risk for motor delays or disabilities. One goal of evaluation is to discriminate infants evidencing motor delays from those developing according to expected norms. Infants weighing less than 1,500 g at birth or experiencing pre-, peri-, or postnatal health complications that result in their spending time in a neonatal intensive care unit are at increased biologic risk for developmental delays, particularly motor delays (Jeng, Yau, Chen, & Hsiao, 2000; Liao & Campbell, 2004; Piper, Pinnell, Darrah, Maguire, & Byrne, 1992). Research indicates that among infants with low birth weight, those who experience both biologic and environmental risk factors are at greatest risk for developmental delays (Blair & Ramey, 1997; Lester & Miller-Loncar, 2000). These children typically are referred as being at dual risk.

Therapists use standardized developmental measures to help make decisions about a child's motor status relative to a normative group. For therapists to make accurate decisions, the measures must be reliable and valid. In the context of classical measurement theory, reliability and validity can vary depending on the characteristics of the children being tested, the context in which a measure is administered, or the individuals who administer the measure. Consequently, published reliability and validity indexes might not be applicable when the child's or examiner's characteristics differ from those reported in test manuals and journal articles (Snyder, Lawson, Thompson, Stricklin, & Sexton, 1993; Thompson, 2003). Reliability and validity, therefore, should be carefully evaluated when selecting measures for use in clinical practice.

Among the standardized measures used by therapists for discriminative purposes with infants at risk for disorders in motor development are the Alberta Infant Motor Scale (AIMS; Piper & Darrah, 1994) and the Peabody Developmental Motor Scale (PDMS). Developed in 1983, the authors of the PDMS revised and renormed the measure in 2000 (PDMS-2; Folio & Fewell, 2000). Although both the AIMS and the PDMS-2 are used to monitor the motor development of infants and to discriminate infants who have motor delays, the theoretical bases of the two measures differ. The authors of the PDMS-2 based their measure primarily on a hierarchical neuromaturational view of motor development. Underlying assumptions of this perspective are that there is a predictable rate and sequence of motor development that reflects maturation of the central nervous system (Case-Smith, 1996). The PDMS-2 includes items for reflex integration and discrete developmental milestones.

In contrast, the authors of the AIMS used a dynamical systems perspective of motor development to guide development of their measure. This perspective posits that motor behaviors emerge as a function of the interaction of several contributing subsystems (e.g., neural, muscular, biomechanical, sensory, cognitive, emotional) in a task-specific context (Jeng et al., 2000). Items on the AIMS reflect a sequence of motor skills with emphasis on observation of components of movement (weight bearing, postural alignment, and antigravity movement) as infants-transition in and out of prone, supine, sitting, and standing.

Several authors have asserted that the features of the AIMS are advantageous for use in clinical practice (Bartlett & Fanning, 2003; Blanchard, Neilan, Busanich, Garavuso, & Klimas, 2004; Jeng, Yau, Chen, & Hsiao, 2000). First, administration time is approximately 20–30 min, which is feasible in the context of high-risk follow-up or evaluation and monitoring programs. Second, items are scored following observation, which is advantageous for infants at risk, who often exhibit signs of stress when they are handled excessively. Third, the emphasis on observing qualitative aspects of movement sensitizes parents and therapists to the components of movement that are useful to support transitions from one position to another (Bartlett & Fanning, 2003).

CONCURRENT VALIDITY OF AIMS AND PDMS SCORES

Several researchers have examined the concurrent validity of AIMS scores with scores from the first edition of the PDMS. Results provide evidence of concurrent validity, although comparisons between AIMS total scale scores and PDMS subscale scores were not consistently reported, particularly for specific age group. The studies are summarized in Table 1. We located no studies that compared the concurrent validity of scores for AIMS and the PDMS-2.

The normative sample for the AIMS consisted of 2,202 full-term infants from Alberta, Canada. The sample was selected from health units throughout the province using proportional random sampling. Stratification variables included age and gender. No information is provided in the manual about the racial or ethnic characteristics of the sample. Infants were born full term and were not at biologic risk. The PDMS-2 normative sample included 2,003 infants and children residing in 46 states and one Canadian province. Sample characteristics, including geographic region, gender, race, rural, or urban residence, ethnicity, family income, parent education, and disability are comparable to statistics for children under 5 years old reported in the 1997 census (U.S. Bureau of the Census, 1997).

INTER-RATER RELIABILITY OF AIMS AND PDMS

Measures of reliability provide information about the potential for measurement error in an observed score. Inter-rater reliability characterizes the consistency between scores obtained independently by two or more raters. Scoring the AIMS requires clinical judgment based on observation of an infant's posture, weight-bearing position, and anti-gravity movement. In contrast, on the PDMS-2, most items are directly administered and scoring criteria are specified. Inter-rater reliability of scores might differ on measures that are completed through observation versus direct administration, particularly for less experienced clinicians or students.

Table 2 summarizes findings research on inter-rater reliability of the AIMS, including sample and examiner characteristics. In general, reliability coefficients were high across examiners who were experienced physical therapists. The majority of reports on inter-rater reliability of the PDMS are for the first edition. Inter-rater reliability of the PDMS-2 was examined by two individuals who scored 60 completed protocols, which were randomly selected from the normative sample. Differences between original scorers

TABLE 1. Summary of Research on Inter-Rater Reliability of the AIMS

Author(s)and date	Sample size	Sample characteristics	Examiner characteristics and training	Findings
Bartlett & Fanning (2003)	12	Preterm infants $<$ 1,500 g *M* age at testing $=$ 8.08 months (SD $=$ 0.41) Ethnicity not specified	Two physical therapists and principal author (each therapist paired with principal author for $n = 7$ and $n = 5$ assessments, respectively) Experience/training not specified	ICC ($n = 7$) $=$ .95 (CI .73 to .99) ICC ($n = 5$) $=$ .98 (CI .87 to .99)
Blanchard, Neilan, Busanich, Garavuso, & Klimas (2004)	6	Six infants reported to be "healthy and at an appropriate birth weight for their gestational age" Infants videotaped monthly from birth through independent walking 14 sessions chosen for interrater reliability coding Ethnicity not specified	Five authors scored 14 videotaped sessions; established "gold standard" scores Eight early intervention providers from CT (2 PT, 3 OT, 1 EI associate) Minimum 3 years in EI No prior experience with AIMS 1.5 hour training session and practice scoring (11 sessions group 1; 4 sessions group 2)	ICCs (2,1) ranged from .82 to .99 across two training groups for four positions and AIMS total score EI examiners agreed with gold standard percentile ranks for 4 of 14 sessions
Darrah, Redfern, Maguire, Beaulne, Watt. (1998)	45	Full-term ($\geq$ 37 weeks) infants from Alberta, Canada Ethnicity not specified	3 pairs of pediatric physical therapists Trained on administration of AIMS (amount and type of training not specified) Monthly calibration sessions during which all 6 therapists assessed infant together and discussed scoring protocol	$r = .99$

(Continue on next page)

TABLE 1. Summary of Research on Inter-Rater Reliability of the AIMS *(Continued)*

Author(s)and date	Sample size	Sample characteristics	Examiner characteristics and training	Findings
Jeng et al. (2000)	45	Preterm infants 1,523 $\pm$ 508 g 31.5 $\pm$ 3.0 weeks gestational age Taiwanese	Six physical therapists (Three for AIMS interrater study) 3–5 years' pediatric experience No prior experience administering AIMS 32-hour training course 90% agreement with expert standard during training	SEM between raters less than 1.3 points ICCs ranged from .73 to .99 across three age groups (0 to 3 months, 4 to 7 months, $>$ 8 months) for four positions and total AIMS score
Piper, Pinnell, Darrah, Maguire, & Byrne (1992)	221	Full-term "normal" infants	Six physical therapists Experience/training not specified Two independent raters scored same infant at same testing period	$r = .996$ to $.998$ Rater 1: $M = 31.36$; SD $= 17.55$ Rater 2: $M = 31.28$; SD $= 17.52$
Piper & Darrah (1994)	253	"Normal" infants recruited through Edmonton Board of Health Birthweight $>$ 2,500 g Gestational age: 38–42 weeks at birth Normal delivery, no abnormality	Six therapists experienced in infant motor assessment Trained in AIMS to $>$ .80 reliability	$r = .9967$ Rater 1: $M = 34.73$; SD $= 18.65$ Rater 2: $M = 34.66$; SD $= 18.63$

TABLE 2. Summary of Research on Concurrent Validity of the AIMS and PDMS

Author(s) and date	Sample size	Sample characteristics	Examiner characteristics	Findings
Piper, Pinnell, Darrah, Maguire, & Byrne (1992)	103	0–13 months of age "Normal infants" from Edmonton Board of Health well-baby clinics	Six physical therapists (working in teams of three) Experience/training not specified	$r = .97$
Piper & Darrah (1994)	68 infants at risk	20 classified as "abnormal"—18 with diagnoses 48 classified as "at risk" based on GA < 32 weeks	Six physical therapists working in teams of three Experience/training not specified	$r = .95$ (at risk and abnormal) $r = .87$ (abnormal) $r = .98$ (at risk)

and the secondary scorers were minimal, with reliability coefficients ranging from .96 to .99 (Folio & Fewell, 2000).

The purpose of this study was to evaluate concurrent validity and inter-rater reliability of the AIMS in a sample of infants at dual risk using the PDMS-2 as the concurrent measure. Four research questions were addressed: (a) How do total scores on the AIMS compare to scores on the reflex, stationary, and locomotion subscales on the PDGMS-2?; (b) Is concurrent validity dependent on age? (Whether infants are $\leq$ 9 months or older than 9 months of age.); (c) How do AIMS and PDGMS-2 scores of trained student examiners compare to scores of experienced therapists?; and (d) Does inter-rater reliability vary as function of whether examiners observed (AIMS) or directly tested (PDGMS-2) motor development?

METHODS

Participants

Participants were 35 infants who were predominantly African American ($n = 34$). Thirty-three of the infants were recruited from a high-risk follow-up clinic located in an urban, public teaching hospital in the southern United States. These 33 infants were Medicaid eligible or were enrolled in the public child health insurance program in the state, indicating that they and their families were living at or below established federal poverty guidelines. The remaining two infants' families had private insurance and received follow-up through a primary care provider in the community. Institutional Review Board associated with the sponsoring university and parents provided informed consent for their infants to participate.

Infants eligible for study participation were those between birth and 18 months chronological age who weighed less than 1,500 g at birth or received care in a neonatal intensive care unit (NICU). Excluded from the study were infants with a diagnosis of cerebral palsy or those with a known chromosomal abnormality or established motor delay or disability.

Staff associated with the high-risk follow-up clinic reviewed the infants' medical records to determine eligibility for participation. Students were permitted to review the infant's medical records to confirm eligibility. If the child met study criteria, one of the experienced examiners approached the parent or legal guardian to explain the study and obtained written informed consent. Forty parents or legal guardians were approached and 35 consented to participate.

Nineteen of the 35 participants were female. Mean birth weight was 1,512 g (SD = 643) and mean gestational age was 30 weeks (SD = 3 weeks). Mean chronological age on the date of study participation was 8 months, 9 days (range = 2 months, 19 days to 16 months, 23 days) and mean corrected chronological age was 5 months, 23 days (range = 9 days to 12 months, 29 days).

Measures

The AIMS consists of 58 dichotomously scored items (*observed* or *not observed*) and is appropriate for administration to infants from birth to 18 months chronological age. Items are organized into four positional subscales: prone ($n = 21$), supine ($n = 9$), sitting ($n = 12$), and standing ($n = 16$). Line drawings of infants in various positions accompany each item in the scoring booklet. Each item is scored based on observation and judgment about three components of motor performance: weight bearing, posture, and anti-gravity movements. The AIMS can be administered and scored in approximately 20–30 min and requires minimal handling of the infant. To score the AIMS, examiners identify the *least* and *most* mature items observed in each of the four positions. These two items constitute the "window" or range of motor skills observed. All items below the least mature item in the window are credited as *observed*. Each observed item is counted and the total raw score is determined by counting the number of observed items in each position. Total raw scores can be converted to percentile ranks that are based on 1-month age-group intervals. Percentile ranks are not provided for the four positions.

The PDGMS-2 gross motor scale consists of 170 gross motor items organized under four subscales: reflexes, stationary, locomotion, and object manipulation. The measure is designed for use with infants and young children from birth to 83 months of age. The reflex subscale is not administered to children older than 12 months while object manipulation is administered only to children older than 12 months. Administration time ranges from 30 to 60 min depending on the child's age. The child is placed in appropriate positions to administer select items (e.g., suspending the child vertically, placing child in kneeling position). Items are scored using a 3-point scale (0 = *unsuccessful*, 1 = *clear resemblance to item criterion but criterion not fully met*, 2 = *successful performance, criterion met*). Basal and ceilings are established for each subscale. To obtain a basal, the child must receive a score of "2" on three consecutive items. A ceiling is obtained when a child scores "0" on three consecutive items. Children receive full credit (2 points) for all items below the basal item. Examiners compute total and subscale raw scores by summing scores obtained on each item. Raw scores are converted to age-equivalents, percentile ranks, subtest standard scores (*mean* = 10, SD = 3), or a composite quotient (*mean* = 100, SD = 15).

Procedures

Before beginning data collection, 21 student examiners received an overview of the theoretical orientations of the AIMS and PDGMS-2 and instruction on administration and scoring procedures by three experienced examiners. The students also reviewed test manuals and published studies on psychometric properties of each measure when used with young children.

All but one student examiner was enrolled in the Master of Physical Therapy program affiliated with an academic health sciences center located in the south central United States. The remaining student examiner was enrolled in a doctoral program in school psychology and was completing a practicum at the same health sciences center. The physical therapy students had limited, if any, experiences with evaluation, assessment, and treatment of young children and had not completed their pediatric coursework in the curriculum. The school psychology student had completed coursework and clinical experiences related to evaluation, assessment, and intervention, but had very limited experiences with infants. All student examiners were Caucasian, female, and they ranged in age from 23 to 27 years.

The three experienced examiners were faculty members affiliated with the academic health sciences center. Examiner 1 had a doctoral degree and certification in early intervention with over 25 years experience working with infants, toddlers, and preschoolers with and without disabilities. Examiner 2 had a doctoral degree in physiology and licensure in physical therapy. At the time of the study, she had coordinated the pediatrics course in the physical therapy curriculum for 3 years. Examiner 3 was a registered occupational therapist, who had a master of health sciences degree with an advanced clinical specialization in pediatrics. She had 8 years of experience in early intervention. All three experienced examiners were Caucasian and female and they ranged in age from 30 to 47 years.

The theoretical overview and instructional session lasted approximately 8 hours. Approximately 5 hours of the session involved lecture, discussion, manual, and score booklet review, and review of procedures for calculating chronological and corrected chronological age. The remaining 3 hours included viewing videotapes demonstrating administration of each measure and scoring both measures from the videotapes. Before the training session, the three experienced examiners scored each videotape independent of the students to determine "gold standard" scores and consensus was reached about scores for each item on each measure. During training, students viewed videotapes in their entirety and then scored each item on the AIMS or PDGMS-2. Student ratings were compared to the scores of the experienced raters and scoring disagreements were discussed. The tapes were reviewed to clarify scoring decisions made by the experienced examiners.

Following the training session, the student examiners administered and scored each test in a clinical setting (e.g., early intervention program or high-risk clinic). The practice tests were given to 12 infants who were typically developing and three infants who were at risk for motor disabilities secondary to low birthweight ($\leq$ 1,500 g). An experienced examiner simultaneously scored the AIMS or PDGMS-2. Scores of student examiners and experienced examiners had to reach a minimum of 85% agreement before the student examiner began data collection.

The order of test administration was determined for each infant by a coin toss, which allowed for counterbalancing of test administration across participants. Two student examiners and one experienced examiner administered either the AIMS or PDGMS-2. The object manipulation subscale of the PDGMS was not administered because the majority of infants ($n = 31$) were below the corrected chronological age of 12 months and the AIMS does not include items for object manipulation. Following administration of the first test and a 15-minute break for the infant, a second set of two student examiners and one experienced examiner administered the second test. On average, each student examiner administered three tests (range one to seven). Parents or legal guardians were present for all sessions. To prevent bias, each group of examiners was unaware of the other group's scores.

Data Analyses

Item scores for both measures were entered into a statistical analysis program. Scores of experienced examiners were used to analyze concurrent validity. Pearson product-moment correlation coefficients between scores for the AIMS and the PDGMS-2 subscales were computed. Correlations were computed for the entire sample and separately for infants whose chronological or corrected chronological age was less than or equal to 9 months ($n = 24$) and infants whose chronological or corrected chronological age was 9 months and older ($n = 11$). We made a decision to evaluate concurrent validity of scores for these two age groups based on results of a study by Liao and Campbell (2004) that suggested that precision of the AIMS decreases after 9 months as there are a fewer number of items.

Scores of student examiners were compared to scores of experienced examiners for analysis of inter-rater reliability. Intra-class correlation coefficients using a two-way random effects ANOVA model were calculated to examine inter-rater reliability and agreement.

RESULTS

The Pearson product-moment correlation coefficients between AIMS scores and PDGMS-2 subscale scores (reflexes, stationary, and locomotion) for the total sample and the two age groups are presented in Table 3. Correlation coefficients generally were high ($r > .80$), positive, and statistically significant ($p \leq .05$). The most noteworthy correlation was between AIMS scores and PDGMS-2 locomotion subscale scores for the total sample ($r = .97$). Correlation coefficients for infants $\leq$ 9 months of age varied

TABLE 3. Pearson Product Moment Correlations Between AIMS and PDGMS-2 Subscale Scores for the Total Sample and Infants $\leq$ 9 Months and $>$ 9 Months of Age

Scale or subscale	AIMS Total	PDGMS-2 Reflex	PDGMS-2 Stationary	PDGMS-2 Locomotion
AIMS	—	0.90	0.92	0.97
PDGMS-2 Reflex	0.82^1 0.78^2	—	0.84	0.90
PDGMS-2 Stationary	0.89^1 0.88^2		—	0.87
PDGMS-2 Locomotion	0.93^1 0.84^2			—

Above the diagonal are correlations for the total sample ($n = 35$).
Below diagonal are correlations for each age group:
^{1} = Infants $\leq$ 9 months of age ($n = 21$).
^{2} = Infants $>$ 9 months of age ($n = 11$).

TABLE 4. Intra-Class Correlation Coefficients (ICC) Between Scores of Experienced and Student Examiners for the AIMS and PDGMS-2 Subscales

Instrument	ICC (95thCI)	*F*(df)	*p*
AIMS	.98 (.96 – .99)	2.07 (2, 52)	.136
PDGMS-2 Reflex Subscale	.98 (.96 – .99)	0.42 (2, 32)	.659
PDGMS-2 Stationary subscale	.99 (.98 – .995)	1.61 (2, 36)	.561
PDGMS-2 Locomotion subscale	1.00 (.99 – 1.00)	0.43 (2, 34)	.654

from .82 to .93. Correlation coefficients for infants > 9 months of age varied from .78 to .88.

The intraclass correlation coefficients (ICC) between scores of experienced and student examiners are presented in Table 4. ICCs were .98 and higher for the AIMS and PDGMS-2 reflex, stationary, and locomotion subscale scores. ANOVA analyses verified high levels of agreement with minimal variance attributed to raters ($p > .05$ for all analyses). Inter-rater reliability of scores did not vary greatly as a function of whether examiners observed (AIMS) or directly tested (PDGMS-2) motor development.

DISCUSSION

The Pearson correlation coefficients are similar to those reported in studies that examined concurrent validity of the first edition of the PDMS and the AIMS (Piper et al., 1992; Piper & Darrah, 1994). The highest correlation was obtained between the PDGMS-2 locomotion subscale scores and AIMS scores. This is a logical finding given the emphasis of the AIMS on observing children as they assume and move in four different positions (prone, supine, sitting, and standing) and the items on the locomotion subscale, which assess the child's movement.

When concurrent validity was analyzed according to whether children were 9 months or younger versus over 9 months of age, the coefficients were somewhat lower for the older infants. This finding might reflect differences in item content on the AIMS versus the PDGMS-2 and that there are fewer items on the AIMS once children pull to stand and begin to walk. In contrast, the PDGMS-2 has many more items above a 9month developmental level. In addition, the PDGMS-2 includes a reflex subscale, and, with the exception of the asymmetrical tonic neck reflex, reflex integration is not measured on the AIMS. However, the moderate correlation (0.78) between AIMS scores and PDMS-2 reflex subscale scores (0.78) for infants older than 9 months most likely occurred because success on AIMS items after 6 months of age requires integration of primitive reflexes. Although not specifically measured on the AIMS, quality of movement may reflect reflex integration.

The intraclass correlations between scores of trained student examiners and expert examiners provide evidence of inter-rater reliability and are comparable to the coefficients reported by Piper et al. (1992, 1993), Darrah et al. (2000), and Jeng et al. (2000), who evaluated the inter-rater reliability of AIMS scores. The intraclass correlation coefficients for the PDMS-2 are consistent with findings reported in the PDMS-2 manual, although the reliability coefficients reported were based solely on scoring similarities rather than direct observations by examiners.

No noteworthy differences in inter-rater reliability were found based on whether raters evaluated infants using the AIMS or the PDGMS-2. In the present study, inter-rater reliability was demonstrated for both the observational assessment (AIMS) and the direct assessment (PDGMS-2). Although the ICC for the total AIMS score was lower than the ICCs for the PDGMS-2 subscales, confidence intervals overlapped for all obtained ICCs. When disagreements on the AIMS occurred, they tended to be those items that either happened quickly (e.g., hands to knees) or were subtle qualitative motor movements difficult for novice examiners to observe (e.g., active versus passive extension, position of arms, and shoulders in prone propping). The slight difference in the number of disagreements across the two measures is not unexpected given the observational structure for completing the AIMS and suggests the importance of training novice examiners before using either of the measures in clinical practice.

Several limitations might have impacted on our findings. First, infants were evaluated with both measures on the same day. Fatigue might have affected performance on the second measure. We attempted to control for this potential confound by randomizing the order of administration and by providing infants a 15-minute rest between tests. Second, infants were evaluated in a clinical setting and not in their natural environments. Factors that might have influenced performance include unfamiliarity of the examiners, room temperature, and unfamiliarity with stimulus used for some items on the PDGMS-2.

The results add to knowledge of reliability and validity of the AIMS and PDMS-2 by providing evidence of concurrent validity of scores for infants at dual risk (biological and environmental) for delays or disorders in motor development. In addition, evidence of inter-rater reliability suggests that, with training, novice examiners can administer an observational measure (the AIMS) reliably.

The AIMS has several features that are advantageous when used with infants, who are at biological risk. First, it requires minimal handling by an unfamiliar examiner. Second, it emphasizes components of motor performance based on contemporary perspectives about motor development and control (Case-Smith, 1996). Finally, administration time for the AIMS is about 20–30 min. Despite these noted advantages, the AIMS might be less appropriate due to the small number of items that discriminate among infants > 9 months of age. The PDMS-2 has more items available to evaluate motor development of older infants and young children providing opportunities to measure change through late infancy and early childhood.

Anecdotally, the student examiners reported that the AIMS also promoted their ability to analyze infant movement, indicating that the AIMS might have good potential as a teaching tool in pediatric physical therapy curricula. Learning about motor development and control using a structured observational measure like the AIMS might be a useful tool for systematic observation of videotapes or observing children in classroom or laboratory settings. Future research might be conducted to explore the utility of the AIMS in a preservice pediatric curriculum.

The results suggest that the AIMS and PDGMS-2 yield valid and reliable scores for infants at dual risk for developmental delay when administered by trained examiners, whether the examiners are novice or experienced clinicians. As the focus in pediatric assessment continues to shift toward naturalistic observations of children's functional ability in everyday settings (Campbell, 2000), additional research about existing standardized motor measures and additional test refinement and development will need to occur. This should include continued emphasis on ensuring all measures used by clinicians have acceptable psychometric properties for children with and without disabilities.

REFERENCES

Bartlett, D. J., & Fanning, J. E. K. (2003). Use of the Alberta Infant Motor Scale to characterize the motor development of infants born preterm at eight months corrected age. *Physical and Occupational Therapy in Pediatrics, 23* (4), 31–45.

Blanchard, Y., Neilan, E., Busanich, J., Garavuso, L., & Klimas, D. (2004). Inter-rater reliability of early intervention providers scoring the Alberta Infant Motor Scale. *Pediatric Physical Therapy, 16,* 13–18.

Campbell, S. (2000). *The child's development of functional movement*. In S. K. Campbell, D. W. Vander Linden, & R. J. Palisano (Eds.), *Physical therapy for children* (2nded., pp. 3– 44). Philadelphia: Saunders.

Case-Smith, J. (1996). Analysis of current motor development theory and recently published infant motor assessments. *Infants and Young Children, 9* (1), 29–41.

Cintas, H. L. (1995). Cross-cultural similarities and differences in development and the impact of parental expectations on motor behavior. *Pediatric Physical Therapy, 7,* 103–111.

Cohen, E., Boettcher, K., Maher, T., Phillips, A., Terrel, L., Nixon-Cave, K., & Shepard, K. (1999). Evaluation of the Peabody Developmental Gross Motor Scales for young children of African American and Hispanic ethnic background. *Pediatric Physical Therapy, 11*, 191–197.

Coster, W. (1995). Critique of the Alberta Infant Motor Scale (AIMS). *Physical and Occupational Therapy in Pediatrics, 15,* 53–69.

Darrah, Redfern, Maguire, Beaulne, & Watt. (1998). Intra-individual stability of rate of gross motor development in full-term infants. *Early Human Development, 52*, 169–179.

Folio, M. R., & Fewell, R. R. (2000). *Peabody developmental motor scales: Examiner's manual*. (2nd ed.). Austin, TX: Pro-Ed.

Individuals with Disabilities Education Act, Amendments of 1997, 20 U.S.C. §1400 *et seq*. (Supp. III 1997). .

Jeng, S .F., Yau, K. I. T., Chen, L. C., & Hsiao, S. F. (2000). Alberta Infant Motor Scale: Reliability and validity when used on preterm infants in Taiwan. *Physical Therapy, 80,* 168–178.

Liao, P. J. M., & Campbell, S. K. (2004). Examination of the item structure of the Alberta Infant Motor Scale. *Pediatric Physical Therapy, 16,* 31–38.

Malina, R. M. (1988). Racial/ethnic variation in the motor development and performance of American children. *Canadian Journal of Sport Science, 13*(2), 136–143.

Piper, M. C., & Darrah, J. (1994*). Motor assessment of the developing infant*. Philadelphia: Saunders.

Piper, M. C., Pinnell, L. E., Darrah, J., Maguire, T., & Byrne, P. J. (1992). Construction and validation of the Alberta Infant Motor Scale (AIMS). *Canadian Journal of Public Health, 83,* (Suppl. 2), S46–S50.

Sameroff, A. J. (1975). Early influences on development: Fact or fancy? *Merrill-Palmer Quarterly of Behavior and Development, 21,* 267–294. .

Snyder, P., Lawson, S., Thompson, B., Stricklin, S., & Sexton, D. (1993). Evaluating the psychometric integrity of instruments used in early intervention research: The Battelle Developmental Inventory. *Topics in Early Childhood Special Education, 13*(2), 216–232.

Thompson, B. (2003). *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage.

U.S. Bureau of the Census. (1997). *Statistical abstract of the United States* (117th ed.). Washington, DC: U.S. Department of Commerce.