Concurrent Validity of the TIMP and the Bayley III Scales at 6 Weeks Corrected Age

Suzann K. Campbell, PT, PhD, **Laura Zawacki, PT, MS, PCS**, **Kristin M. Rankin, PhD**, **Joseph C. Yoder, MS**, **Nicole Shapiro, BA, RN**, **Zhuoying Li, MS**, and **Rosemary White-Traut, RN, PhD**

Department of Physical Therapy, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, Illinois (Dr. Campbell); Department of Physical and Occupational Therapy, University of Illinois Hospital and Health Sciences System, Chicago, Illinois (Ms Zawacki); Department of Women, Children and Family Health Science, College of Nursing, University of Illinois at Chicago, Chicago, Illinois (Mr Yoder, Ms Shapiro, Ms Li and Dr. White-Traut); School of Public Health, University of Illinois at Chicago, Chicago, IL (Dr. Rankin); Children's Hospital of Wisconsin Research Institute, Milwaukee, Wisconsin (Dr. White-Traut).

Abstract

Purpose—Examine agreement between the Test of Infant Motor Performance (TIMP) and the Bayley III.

Methods—One-hundred, forty-five infants born at 29-34 weeks gestation with socioenvironmental risk factors were tested on the TIMP and Bayley III at 6 weeks corrected age (CA). Scores were correlated to assess convergence/divergence of content. Decision analysis using a cutoff of the mean on the Bayley Motor Composite and -.5 and -1 SD from the mean on the TIMP assessed agreement on delay/non-delay.

Results—The TIMP-Bayley Motor Composite correlation was .546, with Cognitive was .310, and with Language was .281. 9% of infants scored below −1.0 SD on the TIMP while no child performed below -1 SD on the Bayley Motor scale (sensitivity 31%).

Conclusions—Convergent validity between the TIMP and the Bayley Motor scale was demonstrated, but no infant showed delay on any Bayley scale. The TIMP is preferred for early assessment of infants.

Keywords

developmental disabilities/diagnosis; infant/premature; infant, newborn; motor skills disorders/diagnosis; neuropsychological tests; predictive value of tests; reproducibility of results; risk; socioeconomic factors

INTRODUCTION

Accurate assessment of the motor performance of infants born prematurely or with other perinatal complications requires use of tests with validity for identifying delay or neurologic impairment. Documentation of the construct validity of a new assessment requires research demonstrating its relationships with other existing tests intended for similar purposes. An exploration of external validity examines the congruence or divergence of the constructs underlying new and criterion tests. Of greater interest to practitioners who use these tools to diagnose delay clinically is an examination of the various cut scores for each tool and the differences in referral decisions based on having selected one tool or the other at a particular age.

The Test of Infant Motor Performance was published in 2001 for use in Neonatal Intensive Care Units (NICUs) and Early Intervention (EI) programs serving infants born prematurely and other infants at risk for developmental delay.[1] The TIMP is a test of functional motor skills with age standards for performance of infants between the ages of 34 weeks postmenstrual age (PMA) and 17 weeks corrected age (CA) based on a normative sample of 990 U.S. infants with a range of medical risk for developmental delay.[2] Previous research demonstrated convergent validity (r = .66) at 3 months CA with the Alberta Infant Motor Scale,[3,4] and validity at 3 months CA for predicting 1) AIMS scores at 12 months CA[5] and 2) Peabody Developmental Motor Scales motor quotients at 4-5 years of age.[6] In a Korean sample TIMP scores from tests performed at term-equivalent age were highly predictive of scores on the Bayley II at 6 months CA,[7,8] but the TIMP has not been compared to the Bayley III.

The Bayley Scales of Infant Development have long been standard tests for measuring motor and cognitive performance of infants.[9] The 3rd edition of the test, the Bayley III,[10] was published in 2006 with new age standards for cognitive, language, and motor development. Separate scales for gross and fine motor performance were developed. Evidence on performance of the Bayley III in clinical situations is accumulating but presents a confusing picture of how this new version of the scales performs when compared with the Bayley II and other tests. On the one hand, Green and colleagues reported clinical results consistent with expected levels of delay at 8-12 months CA in a group of 85 infants born premature.[11] The average score on the motor composite was 94 (SD 17); 22% of the infants had motor composite scores < 85 (1 SD below the normative mean of 100), and 12% had significantly delayed gross motor subscale scores (< 4). On the other hand, Vohr and colleagues found large discrepancies in outcomes of a sample of more than 2600 infants with extremely low birth weight (ELBW) when results from Bayley II assessments at 18-22 months in 2006-2007 were compared with results using the Bayley III on a cohort tested in 2008-2011.[12] The neurodevelopmental impairment rate was reduced from 43% in the earlier period to 13% in 2008-2011. Anderson and colleagues also reported overestimation of performance in infants born full term.[13]

The purpose of this study is to examine the convergent validity of the TIMP and the Bayley III motor scores and the divergent validity of the TIMP and the Bayley III cognitive and language scores at 6 weeks CA in a sample of infants born preterm with dual risk for poor developmental outcome because of social-environmental home conditions. In addition to exploring the correlations among these assessments, we examine the differences in clinical decisions made using threshold cut scores for identification of delayed motor development. Both tests were standardized during approximately the same time period, but given the slightly earlier time for the TIMP (2002-2004)[1] and its more extensive research documenting predictive validity, we use the TIMP as the criterion measure.

METHODS

This report is derived from a larger randomized clinical trial designed to evaluate outcomes of a developmental intervention for infants born preterm. After baseline data collection, 198 infants born at 29-34 weeks gestational age (GA), with at least 2 social-environmental risk factors, were randomly assigned to an attention-only control group or the Hospital-Home Transition: Optimizing Prematures' Environment (H-HOPE) intervention.[14,15] The neurodevelopment of the infants was evaluated at 6 weeks CA via both the TIMP and the Bayley III. No group differences in infant development at 6 weeks CA were identified as a result of the intervention so, for the purpose of this report, we used the entire sample to examine the convergent/divergent validity of the TIMP Z scores and the Bayley III Cognitive, Motor, and Language Composite scores at 6 weeks CA.

Setting

The research was conducted in 2 inner-city Midwestern community hospital NICUs (1 a level II intermediate care unit and the second a level III unit serving infants with special health care needs such as ventilatory support).

Sample

Infants met eligibility criteria if they were born between 29 and 34 weeks GA and had no other major health problems. Infants may have previously received ventilator support or other medical therapies for maintenance (e.g., IV therapy or oxygen therapy via nasal cannula). Their mothers met eligibility criteria if they had at least 2 social-environmental risk factors such as minority status, less than high school education, less than 18 years of age, history of current mental illness (e.g., depression), family income less than 185% of the federal poverty guidelines, more than 1 infant under 24 months, 4 or more infants under 4 years of age in the home, or residing in a disadvantaged neighborhood.

Infant exclusion criteria included the presence of congenital anomalies, necrotizing enterocolitis, brain injury, chronic lung disease, history of prenatal illicit drug exposure or positive toxicology screen. Infants were excluded if their mothers were illicit drug users, not the legal guardian of the infant, or HIV positive. Sample characteristics were obtained via medical record review, with the exception of maternal race which was self-reported by the mother at baseline.

Of the 198 mother-infant dyads enrolled in the intervention study, 149 (75.3%) were retained for the 6-week CA visit, but 3 were not assessed for infant development because the tester was not available at the time of the visit. One additional infant completed the TIMP but not the Bayley III assessment as a result of fatigue. Therefore, 145 infants have scores for the TIMP and the 3 Bayley III scales and are included in this study. Table 1 summarizes the characteristics of the subjects. The average GA of the infants at birth was 32.4 weeks (SD = 1.6). The infants had a moderate degree of medical issues (mean = 70.9) as documented with the Problem-Oriented Perinatal Risk Assessment System (POPRAS) which is used to assign points to a variety of medical conditions.[16,17] Higher scores denote the presence of more medical complications and increased risk for developmental morbidity. At the 6-week CA follow up visit, the mean chronologic age was 13.4 weeks (SD = 1.9).

Procedures—The research was approved by the Institutional Review Boards from the university and the 2 clinical sites. After informed consent was obtained, mothers and infants were randomly assigned to the H-HOPE or the attention-only control group.[15] When the infant reached 32 weeks PMA, the intervention began in the hospital following an initial oral feeding assessment. Infant development was assessed at 6 weeks CA via the TIMP and the

Bayley III. For the 6-week assessments, mothers brought their infants to an examination room in the university's College of Nursing. In most cases the sessions proceeded according to the following schedule: Infants were evaluated first with the TIMP followed by the Bayley. Because this was part of a larger study which also evaluated mother-infant feeding interactions,[15] the mothers were instructed to let the researchers know when they believed that their infants were showing feeding readiness cues. In such instances, the therapist stopped the developmental assessment and resumed after the infants had completed their feedings. The majority of the time (75%) the TIMP and Bayley III were completed in their entirety and in that order before the infants were ready to feed.

Tests

The TIMP is a 42-item assessment of functional gross motor performance.[1] Item responses are related to demands for movement placed on infants in daily life interactions with caregivers,[18] and Rasch psychometric analysis revealed that the items reliably separate infants into 5-6 uniquely different levels of development.[19] All 42 items are administered to infants at 6 weeks CA and raw scores are compared to age norms in 2-week increments. Z scores, percentile ranks and age-equivalent scores can also be obtained.

The Bayley III is used to assess the developmental functioning of infants and toddlers 1 to 42 months of age.[20] The Bayley III consists of 3 administered scales: Cognitive, Language (including receptive and expressive communication subtests), and Motor (including fine and gross motor subtests). The cognitive scale is used to assess sensory-perceptual acuities, discriminations, and the ability to respond to these as well as the early acquisition of object constancy and memory, learning, and problem-solving ability. The language scale is used to evaluate receptive language capabilities, expressive vocalizations and the beginnings of verbal communication. The motor scale provides a means to assess postural control, coordination of the large muscles and finer manipulation skills of the hands and fingers; results can be reported as separate gross and fine motor scores. Scores are presented as raw scores, scaled scores, composite scores, centile ranks, age equivalents, and growth scores.[10,21]

Data Analysis

For this study, we assessed inter-rater reliability for a random 25% of the TIMP and Bayley assessments from a video recording of the original administration of the tests. A physical therapist who is an expert in administering the TIMP re-rated the TIMP, and an expert in administering the Bayley test re-rated the Cognitive, Language and Motor scales of that test. Both were unaware of infant group assessment and scores obtained by the study testers. Inter-rater reliability was determined using the intraclass correlation coefficient (ICC).[22]

Means, standard deviations and proportions were used to describe the sample characteristics, and the TIMP Z scores and Bayley III Composite scores for cognitive, language and motor development were calculated for test comparisons. Pearson product moment correlation coefficients were calculated for the relationship between TIMP and Bayley III Composite scores, and cross-tabulations were generated to contrast the number of infants with indication of delay in motor development according to the TIMP with the number of infants falling below the mean of the Bayley III scores.

To formally compare the difference in clinical decision making regarding motor developmental outcome that would occur with TIMP Z score cutoffs of −0.5, −1.0 and −1.5 SD below the mean and a cutoff at the mean (<100) for the Bayley Motor Composite score, sensitivity, specificity, positive predictive validity (PPV) and negative predictive validity (NPV) were calculated using the TIMP as the criterion measure or reference standard.

Sensitivity was calculated as the proportion of infants with scores below the cutoff on the TIMP Z score that also scored below the cutoff on the Bayley III Motor Composite score. Specificity was calculated as the proportion of infants with scores at or above the TIMP cutoff who also had Bayley III Motor Composite scores above the cutoff. The PPV was calculated as the proportion of infants with a score below the cutoff on the Bayley Motor Composite who also had a score below the cutoff on the TIMP Z score, and the NPV was the proportion of infants with a score at or above the cutoff on the Bayley Motor Composite who also had a TIMP Z score at or above the cutoff.

RESULTS

Rater Reliability

Inter-rater reliability for the TIMP resulted in an ICC of 0.79 (95% CI = 0.60-0.90); for the Bayley Cognitive scale the ICC was 0.73(0.46, 0.86); for the Bayley Language scale the ICC was 0.75 (0.51, 0.87) and for the Bayley Motor scale the ICC was 0.75 (0.46, 0.88).

Test Scores

Table 2 presents the means, standard deviations, and ranges of raw score performance on the TIMP and Bayley III Cognitive, Language and Motor scales at 6 weeks CA, and Table 3 describes the outcomes in standard score terms (Z score for the TIMP and composite scores for the Bayley III). Sixty of the 145 infants (41.4%) scored below the recommended cutoff for delay of a Z score of −0.5 SD on the TIMP.[1] Scores on the Bayley scales were much higher with the means of all composites greater than 100. Ten infants scored below the mean on the cognitive scale (6.9%), 15 (10.3%) on the language scale, and 5 (3.4%) on the motor scale. No infant scored more than 1 SD below the mean on any Bayley III scale. As a result no infants were identified as delayed on the Bayley III scales using the typical cutoff for suspicious performance.

Correlations Among Tests

Table 4 shows the correlations among the TIMP Z scores at 6 weeks and the Bayley III Cognitive, Language, and Motor Scale Composite scores. As expected, the correlation between the TIMP and the Bayley Motor Composite was higher than those between the TIMP and the Bayley Cognitive Composite or Language Composite.

Clinical Decision Comparison

Table 5 presents a 4-fold table comparing the Bayley Motor Composite results split at the mean of the distribution to the TIMP Z score results using a cutoff of −0.5 SD to designate delay. The mean of the Bayley Motor Composite was used because only 5 infants scored below the mean. The sensitivity of the Bayley for agreement with the TIMP in identifying delay is negligible at 8.3% while the specificity is 100% because all infants above the cutoff on the TIMP scored above the mean on the Bayley Motor Composite (Table 6). The PPV is also 100% because the 5 infants scoring below the mean on the Bayley Motor Composite all had delayed scores on the TIMP. The NPV, on the other hand, was poor at 61% because the vast majority of infants tested above average on the Bayley. The overall agreement of the Bayley in reflecting results on the TIMP was 62.1%.

Because the cutoff recommended for identification of delay on the TIMP was derived from data on concurrent and predictive validity for TIMP assessments at 3 months CA, the lower cutoff of −1.5 SD recently published by Korean researchers[8] for comparing TIMP results at term age with Bayley II results at 6 months CA were next examined along with an intermediate cutoff of −1.0 SD. Table 6 shows the results. A cutoff of the mean on the Bayley Motor Composite best matches results using a cutoff on the TIMP Z score of -1.0 SD

with agreement between decisions at 93%, but the sensitivity remains poor at 31%, i.e., only 4 of the 13 infants (31%) identified as delayed on the TIMP have scores below the mean on the Bayley Motor Composite. One infant scoring below the mean on the Bayley achieved a score on the TIMP > 1.0 SD, and 8.96% of the infants were identified as delayed by the TIMP.

In summary, Bayley Motor Composite scores were significantly correlated with TIMP Z scores while Cognitive and Language Composite scores were less strongly related to TIMP performance. No infant was identified as scoring below average on any of the Bayley III scales while the TIMP identified 41% of infants as delayed using a cutoff of -0.5 SD or 9% of infants as delayed using a cutoff of a Z score of −1.0 SD.

DISCUSSION

A comparison of the correlations among the various Bayley scales and the TIMP demonstrated convergence between the Bayley Motor Composite and the TIMP and divergence between the TIMP and the Bayley Cognitive and Language Composites. An examination of the items on each test supports the conclusion that the Bayley Motor Scales (gross and fine) measure a number of the same skills as the TIMP at 6 weeks CA, including head control in upright, prone and supine, head turning and visual following, reaching, and making crawling movements.[1,20] The range of possible raw scores on the Bayley gross and fine motor scales, however, is about 6-15 points, while that of the TIMP at 6 weeks CA is as much as 70-80 points, providing a greater degree of precision along with a wider range of assessed skills, including postural control in supported standing, trunk rotation, and head and trunk control during lateral tipping actions evoking vestibular responses.

The Cognitive and Language scales of the Bayley III consist primarily of items assessing visual and auditory attention and habituation compared to only 4 items on the TIMP assessing attention or visual or auditory search and no habituation items.[1,20] The relatively lower, although still statistically significant, correlations among these scales support a conclusion that they measure divergent constructs with only a small degree of overlap.

Despite a good correlation between the TIMP and the Bayley Motor Composite, the decision analysis shows that use of the 2 tests at 6 weeks CA yields widely divergent results. The Bayley Motor Composite scores averaged 116 and the lowest score obtained by any infant was 94. We are not aware of any research on the predictability of later developmental outcome from Bayley III scores, but based on the results of this study, clearly no infant would be identified as delayed or even suspicious in motor development when using the Bayley Motor scale as the basis for early identification of the need for close surveillance or intervention. On the other hand, using the recommended cutoff of −0.5 SD to identify delay using the TIMP would result in 41.4% of the infants being flagged for close surveillance or referral for intervention, depending on how low the score was. Previous research on the TIMP at 30 days as compared to preschool age outcomes showed an overall accuracy in predicting Peabody Developmental Motor Scale scores < 2 SD from the mean of 80%.[6] The PPV of early scores was 60% such that 40% of infants who scored low at 30 days were found to have normal motor development at 4-5 years of age. PPV improved to 75% and overall accuracy to 87% if testing occurred at 90 days CA. Although one might conclude that 6 weeks CA is too early to identify delay, low scores on the TIMP reflect performance below that of a national sample of 990 infants who were stratified according to perinatal medical risk factors. Thus low scores at 6 weeks CA provide the opportunity to flag infants close to the cutoff or scoring between −.5 and −1.0 SD for intervention to help them close the gap between their performance and that of a national sample from the population of infants born premature for which the TIMP was designed. Vohr and colleagues suggest that

all infants born with ELBW should be offered EI because of their high risk for cerebral palsy and other disabilities.[12] In the case of this sample of infants with moderate biological risk for delay, but socioeonomically disadvantaged backgrounds, however, frequent surveillance and instruction of parents in appropriate activities to do at home are recommended until such time as a more definitive diagnosis can be reached.

The findings of this study add to the accumulating evidence on overestimation of ability when testing using the Bayley III by providing information on a moderate-risk sample of infants born preterm and assessed at 6 weeks CA. Previous studies have shown overestimation of performance at ages from 12 months to 2 years CA for Australian infants born full-term and preterm,[13] infants born with ELBW compared with performance of earlier cohorts on the Bayley II at 18-22 months,[12] infants under 6 months CA enrolled in EI services,[23] infants post-complex cardiac surgery,[24,25] and at 6 months CA English infants born preterm.[26]

Why has this occurred? The most compelling argument for the overestimation of performance being widely reported is that the 2006 norms for the Bayley III scales were developed in a different manner than those for the previous editions of the scales.[24] In an attempt to reflect the broader population of infants in the U.S., the sample for the 2006 norms included more Latino infants and about 10% clinical cases, including infants born preterm and infants with diagnosed disabilities.[10] This sampling decision lowered the overall means of the scaled scores and now boosts the apparent performance of both infants born preterm and infants that are typically developing being compared to the published norms. Moreover, studies have shown that the greatest overestimation of performance is occurring in infants born prematurely.[12,26] As a result, further research on the predictability of scores for later outcomes is critically needed, and studies using the Bayley III scales should always include a control group for comparison with the group of interest.

If the Bayley III is used to make diagnoses, clinicians should be aware that a high cut score should be used. Moore and colleagues suggest a cut score of 80,[26] but our results suggest that any below average score at early ages should trigger parent instruction and close surveillance with repeated assessment. We agree with the suggestion of Moore and colleagues that a consensus statement is needed on the classification of developmental impairment using the Bayley III based on the accumulating research.[26]

Infants with biologic and socioenvironmental risk, such as those in this study, have a high incidence of poor developmental outcomes. For example, Bradley and colleagues found that only 11% of 3-year-old children born prematurely and living in poverty functioned in the normal range in all areas of growth and development.[27] The provision of a home exercise program for infants with low scores on the TIMP at hospital discharge has been successful in significantly raising TIMP scores at 4 months,[28] demonstration of the TIMP for African-American mothers with low income improved their understanding of infant motor development,[29] and predictability to preschool motor development is high by 3 months CA.[6] As a result, we recommend use of the TIMP for early assessment of infants at risk for delayed motor development with the selection of a cut score being based on resources available and the philosophy of EI of the agency.

Conclusion

Although the Bayley III scales have a degree of commonality with the TIMP, no children in this moderate-risk group were identified as delayed by the Bayley III scales at 6 weeks CA. For assessment of motor performance and determination of the need for intervention at early ages in infants at risk, the TIMP is the preferred test.

Acknowledgments

References

1. Campbell, SK. The Test of Infant Motor Performance. Test User's Manual Version 3.0 for the TIMP Version 5. Infant Motor Performance Scales, LLC; Chicago IL: 2012.

2. Campbell SK, Levy P, Zawacki L, Liao PJ. Population-based age standards for interpreting results on the Test of Infant Motor Performance. Pediatr Phys Ther. 2006; 18:119–125. doi: 10.1097/01.pep.0000223108.03305.5d. [PubMed: 16735859]

3. Piper, Mc; Darrah, J. Motor Assessment of the Developing Infant. WB Saunders; Philadelphia, PA: 1994.

4. Campbell SK, Kolobe THA. Concurrent validity of the Test of Infant Motor Performance with the Alberta Infant Motor Scale. Pediatr Phys Ther. 2000; 12:1–8.

5. Campbell SK, Kolobe THA, Wright BD, Linacre JM. Validity of the Test of Infant Motor Performance for prediction of 6-, 9- and 12-month scores on the Alberta Infant Motor Scale. Dev Med Child Neurol. 2002; 44:263–272. doi:10.1111/j.1469-8749.2002.tb00802.x. [PubMed: 11995895]

6. Kolobe THA, Bulanda M, Susman L. Predicting motor outcome at preschool age for infants tested at 7, 30, 60, and 90 days after term age using the Test of Infant Motor Performance. Phys Ther. 2004; 84:1144–1156. [PubMed: 15563255]

7. Bayley, N. Bayley Scales of Infant Development. 2nd ed.. Psychological Corporation; San Antonio: 1993.

8. Kim SA, Lee YJ, Lee YG. Predictive value of Test of Infant Motor Performance for infants based on correlation between TIMP and Bayley Scales of Infant Development. Ann Rehabil Med. 2011; 35:860–866. doi:10.5535/arm.2011.35.6.860. [PubMed: 22506215]

9. Bayley, N. Manual for the Bayley Scales of Infant Development. The Psychological Corporation; San Antonio: 1969.

10. Bayley, N. Bayley Scales of Infant and Toddler Development: Technical Manual. 3rd ed.. Harcourt Assessment; San Antonio: 2006.

11. Green MM, Patra K, Nelson MN, Silvestri JM. Evaluating preterm infants with the Bayley-III: patterns and correlates of development. Res Dev Disabil. 2012; 33:1948–1956. doi:10.1016/j.ridd. 2012.05.024. [PubMed: 22738765]

12. Vohr BR, Stephens BE, Higgins RD, et al. Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. J Pediatr. 2012; 161:222–228. doi:10.1016/j.jpeds. 2012.01.057. [PubMed: 22421261]

13. Anderson PJ, De Luca CR, Hutchinson E, et al. Under-estimation of developmental delay by the new Bayley III Scale. Arch Pediatr Adolesc Med. 2010; 164:352–356. doi:10.1001/archpediatrics. 2010.20. [PubMed: 20368488]

14. Burns K, Cunningham N, White-Traut R, Silvestri J, Nelson MN. Infant stimulation: modification of an intervention based on physiologic and behavioral cues. J Obstet Gynecol Neonatal Nurs. 1994; 23:581–589. doi:10.1111/j.1552-6909.1994.tb01924.x.

15. White-Traut R, Norr K. An ecological model for premature infant feeding. J Obstet Gynecol Neonatal Nurs. 2009; 38:478–490. doi:10.1111/j.1552-6909.2009.01046.x.

16. Davidson, EC.; Hobel, CJ. POPRAS: A Guide to Using the Prenatal, Intrapartum, Postpartum Record. South Bay Regional Perinatal Project Professional Staff Association; Torrence, CA: 1978.

17. Molfese VJ, Thomason B. Optimality versus complications: assessing predictive values of perinatal scales. Child Dev. 1985; 56:810–823. [PubMed: 4042745]

18. Murney ME, Campbell SK. The ecological relevance of the Test of Infant Motor Performance elicited scale items. Phys Ther. 1998; 78:479–489. [PubMed: 9597062]

19. Campbell SK, Wright BD, Linacre JM. Development of a functional movement scale for infants. J Appl Meas. 2002; 3(2):191–205.

20. Bayley, N. Bayley Scales of Infant and Toddler Development: Administration Manual. 3rd ed.. Harcourt Assessment; San Antonio: 2006.

21. Spittle AJ, Doyle LW, Boyd RN. A systematic review of the clinimetric properties of neuromotor assessments for preterm infants during the first year of life. Dev Med Child Neurol. 2008; 50:254–266. doi:10.1111/j.1469-8749.2008.02025.x. [PubMed: 18190538]

22. Fleiss, JL. Statistical methods for rates and proportions. John Wiley & Sons; New York: 1981.

23. Connolly BH, McClune NO, Gatlin R. Concurrent validity of the Bayley-III and the Peabody Developmental Motor Scale-2. Pediatr Phys Ther. 2012; 24:345–352. doi:10.1097/PEP. 0b013e318267c5cf. [PubMed: 22965209]

24. Acton BV, Biggs WSG, Creighton DE, et al. Overestimating neurodevelopment using the Bayley-III after early complex cardiac surgery. Pediatrics. 2011; 128:e794–e800. doi:10.1542/peds. 2011-0331. [PubMed: 21949148]

25. Long SH, Galea MP, Eldridge BJ, Harris SR. Performance of 2-year-old children after early surgery for congenital heart disease on the Bayley Scales of Infant and Toddler Development. Early Hum Dev (Third Edition.). 2012; 88:603–607. doi:10.1016/j.earlhumdev.2012.01.007. [PubMed: 22336496]

26. Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. J Pediatr. 2012; 160:553–558. doi:10.1016/j.jpeds.2011.09.047. [PubMed: 22048046]

27. Bradley RH, Whiteside L, Mundfrom DJ, et al. Early indications of resilience and their relation to experiences in the home environments of low birthweight, premature children living in poverty. Child Dev. 1994; 65:346–360. [PubMed: 8013226]

28. Lekskulchai R, Cole J. Effect of a developmental program on motor performance in infants born preterm. Aust J Physiother. 2001; 47:169–176. [PubMed: 11552873]

29. Goldstein LA, Campbell SK. Effectiveness of the Test of Infant Motor Performance as an educational tool for mothers. Pediatr Phys Ther. 2008; 20:152–159. doi:10.1097/PEP. 0b013e3181729de8. [PubMed: 18480714]

Table 1

Infant Characteristics (n = 145)

	Mean or %	SD	Minimum	Maximum
Male/Female Sex (%)	52 / 48			
Latina/African-American Maternal Race/Ethnicity (%)	50/50			
Multiple/Singleton Birth (%)	10 / 90			
Gestational Age	32.4	1.6	29	34
Birth Weight	1834	409	1000	3146
Apgar Score 5 min	8.3	1.09	4	9
Inf Morbidity Score (POPRAS)	70.9	19.8	36	136
Chronological age at 6 week CA follow-up visit, weeks	13.4	1.9	10	18

Table 2

Descriptive Statistics for Raw Scores for TIMP and Bayley III Subscales at 6 Weeks Corrected Age (N = 145)

	Mean	SD	Minimum	Maximum
TIMP	78.7	9.1	56	104
Bayley Cognitive	9.52	2.15	5	24
Bayley Language				
Receptive Communication	5.84	2.47	4	33
Expressive Communication	4.66	1.11	2	8
Bayley Motor				
Fine Motor	6.27	1.26	3	9
Gross Motor	10.35	2.61	4	16

Table 3

Descriptive Statistics for Performance on the TIMP and Bayley at 6 Weeks Corrected Age (N=145)

	Mean	Standard Deviation	Minimum	Maximum
TIMP Z Score	−0.33	.55	−1.60	1.12
Bayley Cognitive	109	8	85	135
Bayley Language	108	8	94	129
Bayley Motor	116	8	94	136

Table 4

Correlations Among TIMP and Bayley Standard Scores at 6 Weeks Corrected Age (N = 145)

	Bayley Cognitive Composite	Bayley Language Composite	Bayley Motor Composite
TIMP Z score			
Pearson correlation	.310	.281	.546
Significance	<.0001	.001	<.0001
Bayley Cognitive Composite			
Pearson correlation		.236	.303
Significance		.004	<.0001
Bayley Language Composite			
Pearson correlation			.245
Significance			.003

Table 5

Comparison of Clinical Decisions for TIMP Cutoff of −0.5 SD and Bayley Motor Composite Cutoff of the Mean

Test	TIMP <−0.5 SD	TIMP −0.5 SD	Totals
Bayley < 100	5	0	5
Bayley 100	55	85	140
Totals	60	85	145

Table 6

Comparison of Clinical Decisions for TIMP Cutoffs of −0.5, −1.0 and −1.5 SD and Bayley Motor Composite Cutoff at the Mean

Cutoff on TIMP	Sensitivity	Specificity	Positive Predictive Validity	Negative Predictive Validity
<−0.5	.08	1.00	1.00	.61
<−1.0	.31	0.99	0.80	.94
<−1.5	.00	0.97	0.00	.99