Published in final edited form as:

Acta Paediatr. 2012 February ; 101(2): e55–e58. doi:10.1111/j.1651-2227.2011.02517.x.

Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: Are we measuring the same thing?

Jean R Lowe, PhD¹, Sarah J Erickson, PhD², Ron Schrader, PhD³, and Andrea F Duncan, MD¹

¹University of New Mexico School of Medicine

²University of New Mexico Department of Psychology

³University of New Mexico Clinical and Translational Science Center

Abstract

Aim—To better understand differences between Bayley Scales 3rd edition (Bayley III) Cognitive Scale and Bayley Scales 2nd edition (Bayley II) Mental Developmental Index (MDI) in 18–22 month old children born term and preterm; and to create a conversion algorithm using Bayley II MDI to calculate Bayley III Cognitive score.

Methods—This study included 51 term and 26 preterm children between 18 and 22 months, ages adjusted for prematurity. Children's scores on Bayley II MDI and Bayley III Cognitive Scale were compared using t-tests. The items from Bayley II MDI were used to calculate a score for the Bayley III Cognitive Scale. ANCOVA was used to create a conversion scale.

Results—Bayley III Cognitive scores were significantly higher than Bayley II MDI scores for term and preterm toddlers combined and separately ($p < .0001$). A conversion formula to convert Bayley II MDI to a Bayley III Cognitive score was calculated.

Conclusion—Term and preterm children had similarly elevated scores on the Bayley III calculated Cognitive score compared to the previous Bayley II MDI score. The use of a conversion algorithm maybe helpful in studies that used both Bayley editions in order to get comparable outcome measurements within a clinical or research paradigm.

Keywords

Developmental outcome; cognition development; prematurity

The Bayley Scales of Infant and Toddler Development (Bayley),¹ currently one of the most widely used developmental assessment tools in preschool aged children, is considered to be an integrative developmental assessment that borrows from different areas of child development research, including the Gesell scale.^{2,3} The Bayley is now being used extensively in both clinical and research paradigms to diagnose developmental delays and to determine subsequent qualification for early intervention services.⁴ For research purposes, tools such as the Bayley are used to assess developmental outcome related to medical diagnoses such as prematurity,⁵ drug exposure⁶ and genetic disorders.^{7,8}

Prior to 2006, the Bayley Scales of Infant and Toddler Development 2nd edition (Bayley II)⁹ was the standard tool for assessing outcomes among high risk infants and young children.

The Bayley II Mental Developmental Index (MDI) was designed to assess cognition through evaluation of sensory-perception, knowledge, memory, problem solving, and early language. The Bayley II MDI thus measures a combination of early cognitive and language development. This heterogeneous content domain is a great limitation of the Bayley II, as a low MDI score could represent a delay in language skills, cognition, or both.¹⁰ In 2006, the newly revised Bayley Scales of Infant and Toddler Development 3rd edition (Bayley III)¹¹ was introduced. The Bayley III Cognitive Scale was created in an attempt to isolate cognitive skills from language skills. After the introduction of the Bayley III, researchers found that rates of developmental delay were greatly decreased using the Bayley III¹⁰ compared to those found using the Bayley II, and became concerned regarding the effects of this discrepancy on interpretation of results obtained from longitudinal studies incorporating both scales. In addition, the validity of previously published articles that addressed developmental outcomes for extremely preterm children using the Bayley II were questioned.¹² The unanswered question remains: how much of this controversy is due to the different versions of the test¹³ and what strategies can be used in current research to allow comparison of scores obtained using both editions of the Bayley Scales so as to allow interpretation of extensive data collected using the Bayley II?

The objectives of this study were thus to: 1- better understand the test score differences between the Bayley II MDI and Bayley III Cognitive scores for both term and preterm children, and 2- create a conversion algorithm using the Bayley II MDI scores to calculate corresponding Bayley III Cognitive scores to aid in research that includes both versions of the Bayley Scales.

METHODS

Subjects

This study consisted of 77 children born term (N = 51) and preterm (N = 26) between the ages 18 and 22 months (age adjusted for prematurity). Testing occurred between January 2001 and September 2005, which was prior to the development of the Bayley III. Child ethnicity was determined by maternal report and was representative of our state including White, Black, Hispanic and Native American children. This study was approved by the Institutional Review Board Approval. Permission for the study was obtained by the legal guardian when the child was tested.

Term infants were recruited from the University of New Mexico Pediatric Well Baby Clinics. All preterm participants were recruited from the University Children's Hospital Neonatal Intensive Care Unit Follow-up Clinic. In order to be considered pre-term, infants had to have a birth weight < 1500 grams and have a gestational age <32 weeks gestation. Toddlers were excluded from the study if they had been prenatally exposed to drugs, were visually/hearing impaired, had a known genetic abnormality, and/or did not reside with their biological families. Conversion of the Bayley II Mental Developmental Index into a Bayley III Cognitive Score

There are 48 items in the 14 to 22 month window of the Bayley II MDI domain and 29 of them correspond exactly to items included on the Bayley III Cognitive Scale. Of the 32 items in the 14 to 22 month range on the Bayley III Cognitive Scale, only 3 of the test items were not on the earlier Bayley II version (play with toys, ball, and ice-cream puzzle). We used the larger 14–22 month window of items to re-score test items from the Bayley II MDI to the Bayley III Cognitive Scale, as many preterm children tested at 18 to 22 months are delayed and require the use of test items appropriate for younger children in order to determine their abilities. Thus, twenty-nine of the 48 items in the 14 to 22 month window of the Bayley II MDI were used to rescore the Bayley III Cognitive Scale. A review of the 48

items in the Bayley II MDI 14–22 month window indicated that 37% were cognitive items, 29% were expressive language items, 14% were receptive language items, 14% were fine motor items and 6% were items no longer used.¹⁴

A rule was established wherein, if a missing item was within the child's basal score for that domain (i.e., the child had correctly answered all other items), they were given credit for the item; if the item was within the ceiling score (i.e. the child failed five sequential items) they were not given credit. If the item was within the child's "middle range" (had both passes and fails on items), credit was not given for the missing items. This rule was determined a priori and used consistently for all the children in the study. We then obtained a raw score from the correct items and recalculated the Bayley III Cognitive score.

Statistical Analysis

In order to construct a conversion from Bayley II MDI score to Bayley III Cognitive Scale, Analysis of Covariance (ANCOVA) was performed with Bayley III Cognitive score as the dependent variable; term/preterm, gender, and primary language as categorical independent variables; and the Bayley II MDI score as a continuous independent variable. All two-way interactions were allowed, and non-significant terms (as defined by F-test p-value < 0.05) were removed by backward elimination. T-tests were used to compare the demographics and Bayley scores for the preterm and term groups. The corresponding author takes responsibility for vouching for the integrity of the data and the accuracy of the data analysis.

RESULTS

Preliminary analyses revealed that of the 77 children included (51 term and 26 preterm), there were significantly more males, Hispanic children and children with Spanish as a primary language in the term group compared to the preterm group. The two groups were not significantly different on income, maternal education level or age at testing (Table 1). As expected, the preterm group had significantly lower birth weight and gestational age. In the preterm group, 88% were diagnosed with bronchopulmonary dysplasia; 33 % had mild intraventricular hemorrhage, and 1 child (4%) had grade 4 intraventricular hemorrhage. The remaining 63% had no intraventricular hemorrhage.

T-tests indicated that the term children had significantly higher standard scores, on average, than the preterm children on both the Bayley II MDI ($p=0.003$) and the Bayley III Cognitive Scale ($p=0.005$). The standard score discrepancy between the Bayley II and Bayley III scores for the term children (discrepancy score of 14.9) and preterm children (discrepancy score of 18.1) was not significantly different ($p=0.14$). The ANCOVA revealed no interactions or main effects that were statistically significant except for the Bayley II MDI score ($p < 0.0001$), which allowed the groups to be summarized by a single conversion equation. The predicted values from the ANCOVA are labeled as the *Least squares line* in Figure 1.

When both preterm and term groups were combined, the Bayley II MDI and Bayley III Cognitive scores were significantly different ($p<.0001$; $R^2=0.61$), with greater differences for the lower MDI scores (see Figure 1). The conversion from the Bayley II MDI score to the Bayley III Cognitive score is represented by the least squares line in Figure 1, where the Bayley III Cognitive score is computed as 59% of the Bayley II MDI score plus 52. This formula is the regression equation from the ANCOVA. The formula is provided in Figure 1 and Table 2 provides converted scores derived from the formula.

DISCUSSION

Our finding that Bayley III Cognitive scores are significantly higher than Bayley II MDI scores for both term and preterm children is consistent with findings by Anderson and colleagues.¹⁰ The test score discrepancy between the two versions of the test was similar for the term and preterm groups, indicating that both the term and preterm children have similarly elevated scores on the Bayley III Cognitive Scale compared to the previous Bayley II MDI score. The test score differences between the Bayley II MDI and the Bayley III Cognitive Scale were most extreme at the lower end of the scale, as a child who previously had an MDI score of 60 would obtain a Cognitive score of 87 on the newly revised Bayley III. It is therefore evident that simply substituting the Bayley III Cognitive score for the Bayley II MDI score is not valid in cohort studies that include results gained using both versions of the Bayley Scales, and using an algorithm may be preferable. In addition, our algorithm suggests that simply adding a constant to Bayley II MDI scores to obtain Bayley III Cognitive scores is not appropriate. Another possibility to address the Bayley discrepancy would be to use the Bayley III Language scale as Robertson and associates¹³ did not find a significant difference between the Bayley II MDI and the Bayley III Language Composite score.

The Bayley II and Bayley III normative samples are representative of the U.S. population in 1988 and 2000, respectively. The changes in demographic characteristics, particularly parent education level and the ethnic composition of the U.S. population, could contribute to the differences found in the MDI and Cognitive scores.¹⁴ Previous studies have found greater than 7 points difference between the Bayley II MDI and Bayley III Cognitive score, especially for children with medical problems such as those born preterm¹⁵ or after cardiac surgery.¹⁶

As Bayley scores are often utilized to determine eligibility for early intervention services, it is possible that the relative inflation of the Bayley III Cognitive score may result in previously eligible children no longer qualifying for early intervention services. This is especially troubling because the largest discrepancy was seen at the lowest end of the scores, representing the children who may benefit most from early intervention services,¹⁷ but now have scores that would make them appear to not need services.

The finding that the Bayley III yields higher Cognitive scores calls into question the validity of previous findings of morbidity using older versions of the Bayley Scales;^{12,15} as well as the validity of current findings using the newer version.¹⁶ Research findings guide provision of education for families and decision-making in regard to medical conditions such as extreme prematurity.^{4,5} Because there are few developmental tests available for infants and toddlers, both researchers and clinicians rely heavily on the Bayley Scales. Thus it is extremely important to understand and address the discrepancies between MDI and Cognitive scores.

The method of comparison of the Bayley II and Bayley III that we used in the current study was different from that used by Anderson¹⁰ and Robertson,¹³ as we took the previously administered Bayley II and used those responses to re-score the Bayley III. There are benefits and limitations to this method. One benefit is that we avoided a fatigue effect that may be evident when testing time is extended, as is the case when both versions of the Bayley are administered sequentially. As our results are very similar to those found by Anderson¹¹ and Robertson,¹⁴ we infer that both methods are useful, and ours may be preferable when previous test protocols are available, or if time to administer additional test items is not feasible. A limitation to our study is that our preterm group was small, although we were able to increase our power by combining the term and preterm group for the

regression analysis. The combination of these two groups was justified as our ANCOVA results indicated that there were no statistical differences in the score discrepancies found between the two versions of the Bayley due to prematurity or gender.

In conclusion, this study found that the Bayley II MDI and the Bayley III Cognitive scales yield significantly different scores. It is imperative that we consider the consequences of using test scores from changing versions of a test as the primary outcome measure in research studies. This is especially true when research methods include scores obtained from two versions of this test, as is the case when a single study spans the time frame during which the Bayley II was replaced with the Bayley III. We have provided a conversion algorithm as one alternative when trying to use both the Bayley II and Bayley III results in research studies. More generally, we must be aware of the limitations of using a single measure to capture the richness of development.

Acknowledgments

Statistical analysis was supported by a grant from the University of New Mexico Clinical Translational Science Center (1UL1RR031977-01). The assessments were supported by a grant from the University of New Mexico Pediatric Research Committee. We thank the parents and their children that participated in our study.

ABBREVIATIONS

Bayley II	Bayley Scales of Infant and Toddler Development – 2 nd edition
Bayley III	Bayley Scales of Infant and Toddler Development – 3rd edition
Bayley	Bayley Scales of Infant and Toddler Development
MDI	Mental Developmental Index
ANCOVA	Analysis of Covariance

References

1. Bayley, N. Bayley Scales of Infant Development Manual. San Antonio, TX: The Psychological Corporation; 1969.
2. Gesell, AL.; Halverson, HM.; Amatruda, CS. The first five years of life: A guide to the study of the preschool child, from the Yale Clinic of Child Development. New York: Harper & Brothers; 1940.
3. Aylward GP. Conceptual issues in developmental screening and assessment. J Dev Behav Pediatr. 1997; 18:340–349. [PubMed: 9349978]
4. Aylward GP. Developmental screening and assessment: what are we thinking? J Dev Behav Pediatr. 2009; 30(2):169–173. [PubMed: 19363370]
5. Stoll B, Hansen NI, Adams-Chapman I, Fanaroff AA, Hintz SR, Vohr B, et al. Neonatal outcome of extremely preterm infants from the NICHD Neonatal Research Network. Pediatrics. 2010; 126:443–456. [PubMed: 20732945]
6. Messinger DD, Bauer CR, Das A, Seifer R, Lester BM, Lagasse L, et al. The maternal lifestyle study: Cognitive, motor and behavioral outcomes of cocaine-exposed and opiate-exposed infants through three years of age. Pediatrics. 2004; 113:1677–1685. [PubMed: 15173491]
7. Peters SM, Goddard-Finegold J, Beaudet AL, Madduri N, Turcich M, Bacino CA. Cognitive and adaptive behavior profiles of children with Angelman syndrome. American J of Med Genetics. 2004; 128:110–113.
8. Most DE, Fidler DJ, Laforce-Booth C, Kelly J. Stress trajectories in mothers of young children with Down syndrome. J Intell Disabil Res. 2006; 50:501–514.
9. Bayley, N. Bayley Scales of Infant Development Manual. 2. San Antonio, TX: The Psychological Corporation; 1993.

10. Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW. Underestimation of Developmental Delay by the New Bayley 3rd edition Scale. *Arch Pediatr Adolesc Med.* 2010; 16(4):352–356. [PubMed: 20368488]
11. Bayley, N. Bayley Scales of Infant Development Manual. 3. Antonio, TX: The Psychological Corporation; 2006.
12. Vohr BR, Wright LL, Poole WK, McDonald SA. Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks' gestation between 1993 and 1998. *Pediatrics.* 2005; 116 (3):635–643. [PubMed: 16143580]
13. Robertson CM, Hendson L, Biggs W, Acton B. Application of the Flynn effect for the Bayley III Scales. *Arch Pediatr Adolesc Med.* 2010; 164(4):1072–1073. [PubMed: 21041604]
14. Bayley, N. Bayley Scales of Infant Development third edition: Technical Manual. Antonio, TX: The Psychological Corporation; 2006.
15. Vohr BR, Wright LL, Dusick AM, Perritt R, Poole WK, Tyson JE, et al. Center differences and outcomes of extremely low birth weight infants. *Pediatrics.* 2004; 113(4):781–789. [PubMed: 15060228]. [PubMed: 15060228]
16. Acton BV, Biggs WS, Creighton DE, Penner KA, Switzer HN, Thomas JH, et al. Overestimating neurodevelopment using the Bayley-III after early complex cardiac surgery. *Pediatrics.* 2011; 128e (794):2011–0331.
17. Bear LM. Early identification of infants at risk for developmental disabilities. *Pediatr Clin N Am.* 2004; 51:685–701.

Key Notes

The calculated Bayley III Cognitive scores were significantly higher for both preterm and term 18 – 22 month olds compared to Bayley II MDI scores. Because this has serious implications, a conversion formula has been provided.

Figure 1.

Scatter plot of Bayley-III Cognitive Score vs. Bayley-II Mental Developmental Index showing pre-term and term children along with the least squares regression line. The equation of the least squares line (i.e. the conversion formula from Bayley-II to Bayley-III) is: Bayley-III score = (0.59 X Bayley III) + 52.

Table I

Demographics and Testing Information

	Term	Preterm
Birth weight (grams) *	3365±486	989±234
Gestational Age (wks) *	39.3±1.5	27.5±1.9
Days on Ventilation	0	23.1±26.6
Gender: Male	30 60%	11 44%
Maternal Age (yr)	27.2±6.2	30.5±9.2
Child Age at Testing (months)	19.2±1.5	19.7±1.4
Maternal Education # (%)		
<High School	29 (57)	12 (48)
High School	12 (23)	9 (32)
>High School	10 (20)	5 (20)
Income ⁺	2.0±2.2	2.5±2.2
Non-English Primary Language # (%)	14 (26)	3 (12)
Bayley-II MDI score *	91.0±11.3	80.3±14.6
Bayley – III Cognitive score *	105.9±8.6	98.4±11.1

All values presented in means ±standard deviation unless otherwise indicated

⁺ 1.<10,000; 2. 10,000–20,000 3. 20,000–30,000 4. 40,000–50,000 5. >50,000

* p<0.05

Table 2

Conversion table for Bayley II MDI scores to Bayley III Cognitive scores

Bayley II MDI	60	65	70	75	80	85	90	95	100	105	110
Bayley III Equivalent	87	90	93	96	99	102	105	108	111	114	117