

Bayley-III Scales of Infant and Toddler Development: Transcultural Adaptation and Psychometric Properties¹

Vanessa Madaschi
*Universidade Presbiteriana Mackenzie,
São Paulo-SP, Brazil*

Elizeu Coutinho Macedo
*Universidade Presbiteriana Mackenzie,
São Paulo-SP, Brazil*

Tatiana Pontrelli Mecca
*Centro Universitário FIEO,
São Paulo-SP, Brazil*

Cristiane Silvestre Paula²
*Universidade Presbiteriana Mackenzie,
São Paulo-SP, Brazil*

Abstract: Scales with evidence of validity and reliability are important to evaluate child development. In Brazil, there is a lack of standardized instruments to evaluate young children. This study investigated the psychometric properties of the Bayley Scales of Infant Development, Third Edition (Bayley-III). It was translated into Brazilian Portuguese, culturally adapted and tested on 207 children (12-42 months of age). Evidence of convergent validity was obtained from correlations of the Bayley-III with the: Peabody Developmental Motor Scale 2, Leiter International Performance Scale-R, Expressive Vocabulary Assessment List and Peabody Picture Vocabulary Test. Exploratory factor analyses showed a single component explaining 86% of the variance, supported by goodness-of-fit indexes in confirmatory factor analysis. The Bayley-III demonstrated good internal consistency with alpha coefficients greater than or equal to .90 and stability for fine motor scale only. These robust psychometric properties support the use of this tool in future national studies on child development.

Keywords: childhood development, psychometrics, intellectual development, language, motor skills

Escalas Bayley-III de Desenvolvimento Infantil: Adaptação Transcultural e Propriedades Psicométricas

Resumo: Escalas com evidências de validade e precisão são importantes para avaliação do desenvolvimento infantil. No Brasil, há escassez de instrumentos padronizados e normatizados para a primeira infância. Este estudo investigou as propriedades psicométricas da Bayley Scales of Infant and Toddler Development, terceira edição (Bayley-III) que foi traduzida, adaptada para o português e testada com 207 crianças (12-42 meses). Evidências de validade convergente foram observadas entre a Bayley-III e: Peabody Developmental Motor Scale-2, Escala Internacional de Inteligência Leiter-R, Lista de Avaliação de Vocabulário Expressivo e Teste de Vocabulário por Imagens-Peabody. Análise fatorial exploratória indicou componente que explica 86% da variância, corroborado por bons índices de ajustes na análise fatorial confirmatória. A Bayley-III apresentou boa consistência interna com coeficientes alfa a partir de 0,90 e boa estabilidade teste-reteste apenas para a escala motora fina. Estas adequadas propriedades psicométricas podem contribuir para o avanço nas pesquisas em contexto nacional na área de avaliação do desenvolvimento infantil.

Palavras-chave: desenvolvimento infantil, psicometria, desenvolvimento cognitivo, linguagem, habilidades motoras

Escalas de Desarrollo Infantil Bayley-III: Adaptación Transcultural y Propiedades Psicométricas

Resumen: Escalas con evidencia de la validez/fiabilidad son importantes para la evaluación del desarrollo infantil. En Brasil, faltan instrumentos estandarizados/normalizados para la evaluación en la primera infancia. Este estudio investigó las propiedades psicométricas de las Bayley Scales of Infant and Toddler Development-III, traducida y adaptada al portugués. Se evaluaron 207 niños (12-42 meses). Evidencias de validez convergente se observaron entre Bayley-III con: Escala de Desarrollo Motor de Peabody 2, Escala de Inteligencia Internacional Leiter-R, Lista de Evaluación de Vocabulario y Prueba de Vocabulario Expresivo Imágenes Peabody. El análisis factorial exploratorio indicó un componente que explica el 86% de la varianza, corroborado por buenos índices de ajuste en el análisis factorial confirmatorio. Bayley-III mostró buena consistencia interna, con coeficientes alfa de 0,90. La adecuación de las propiedades psicométricas puede contribuir al avance de la investigación en el contexto nacional en el área de evaluación del desarrollo infantil.

Palabras clave: desarrollo infantil, psicometría, desarrollo cognitivo, lenguaje, destreza motora

¹ Paper derived from the first author's master's thesis under supervision of the fourth author, defended in 2012, at the Graduate Program in Developmental Disorders of the Universidade Presbiteriana Mackenzie.
Support: Coordination for the Improvement of Higher Education Personnel (CAPES/Mackenzie-IPM, Cod Mat 71051791).

² Correspondence address:
Cristiane Silvestre Paula. Universidade Presbiteriana Mackenzie, Programa de Pós-Graduação em Distúrbios do Desenvolvimento. Rua da Consolação, 930, prédio 28. CEP 01302-000. São Paulo-SP, Brazil. E-mail: csilvestrep09@gmail.com

Reliable scales with evidence of validity and reliability are important for the clinical investigation of early developmental delays (Santos, Araújo, & Porto, 2008). In Brazil, the challenge of identifying developmental disabilities in young children is worsened by the lack of standardized instruments. One of the only validated tools available for the assessment of child development in Brazilian Portuguese is the Escala de Desenvolvimento do Comportamento da Criança no Primeiro Ano de Vida (Pinto, Vilanova, & Vieira, 1997). However, this instrument only assesses motor and communication development and is restricted to the first 12 months of life; therefore it cannot be used in toddlers or in longitudinal studies.

The Bayley Scales of Infant Development, currently in its third edition (Bayley-III), is internationally recognized as one of the most comprehensive tools for the assessment of young children. It is widely used in research, in clinical practice, and to evaluate interventions, because it assesses several developmental domains and has a solid theoretical background with robust psychometric properties (Bayley, 2006). Although the Bayley-III has been used to assess child development in many countries, recent studies have shown that it tends to estimate differently children with typical development and at risk for developmental delay depending on geographic location (Acton et al., 2011; Milne, McDonald, & Comino, 2012; Moore, Johnson, Haider, Hennessy, & Marlow, 2012; Reuner, Fields, Wittke, Löprrich, & Pietz, 2013; Yu et al., 2013). Consequently, the use of the original American Bayley scale without adaptations is not recommended, because economic, ethnic and cultural factors can lead to the incorrect assessment of developmental delays (Fleuren, Smit, Stijnen, & Hartman, 2007).

In the last five years, there have been several publications using the Bayley scales to assess developmental delays in Brazilian children (Eickmann, Malkes, & Lima, 2012; Fernandes et al., 2012; Ferreira, Melo, & Silva, 2014; Hentges et al., 2014; Silveira & Enumo, 2012). However, there are no studies about the translation and transcultural adaptation of the Brazilian Portuguese version of the Bayley-III scale, or its psychometric properties. There is a single study about the Bayley Infant Neurodevelopment Screener for children aged 12-24 months (Guedes, Primi, & Kopelman, 2011). Therefore, studies on these topics are still necessary.

Due to the importance of this instrument in assessing child development, the objectives of this study were to translate, culturally adapt and validate the Brazilian version of the Bayley-III in a sample of children in daycare centers in a city in the greater São Paulo area. Specifically, this study aimed to investigate the internal consistency and item homogeneity as well as evidence of validity based on internal structure and in relation to external variables.

Method

We first obtained formal permission to translate and validate the Bayley-III scale from the American publishers of this tool (NCS Pearson). We then started the process of developing a Brazilian version of Bayley-III, following the

recommendations of Hambleton and Patsula (1999) and Herdman, Fox-Rushby and Badia (1998) for translation and adaptation of a test, considering conceptual, item, semantic, operational, measurement and functional equivalences. Each step of the process will be presented in the results section.

Participants

Barueri is a city with approximately 260 thousand inhabitants located in the metropolitan region of São Paulo. For data collection, we selected two out of the 21 daycare centers in the city. There were 350 children aged 12-42 months registered in the two selected centers. Children who were born at term, without any chronic diseases and known developmental disorders were eligible for inclusion. Three children were excluded: one with autism spectrum disorder and two with cerebral palsy. From the total eligible families of 347, 101 refused to participate (sample loss of 29.1%), and from the remaining 246, we randomly recruited 207 children (49.27% girls) aged 11 to 42 months to include in the study. They were distributed according to the categories proposed in the Bayley-III technical manual: 9 children aged 12 months to 13 months and 15 days; 9 children aged 13 months and 16 days to 16 months and 15 days; 9 children from 16 months and 16 days to 19 months and 15 days; 9 children from 19 months and 16 days to 22 months and 15 days; 33 children from 22 months and 16 days to 25 months and 15 days; 34 children from 25 months and 16 days to 28 months and 15 days; 34 children from 28 months and 16 days to 32 months and 15 days; 35 children from 33 months and 15 days to 38 months and 15 days; 35 children from 39 months and 15 days to 42 months and 15 days.

All the children attended the daycare center full time, most of them were Caucasian (74%), belonging to families with the following income: 28% receiving between 1-2 times minimum monthly wage, 56% with 3-4, and only 16% higher than that. The majority of mothers (58%) and fathers (53%) had completed high school or had a lower level of study (19% of mothers and 27% of fathers).

Ten out of the 207 children (4 children 12-24 months of age, and 6 children 25-42 months of age), half boys and half girls, also participated in the test-retest reliability study. All of them were first evaluated by one expert, and 15 days later, by another to avoid memory contamination and contamination among evaluators.

Instruments

The *Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III)* is an individually administered scale that assesses five key developmental domains in children between 1-42 months of age: cognition, language (receptive and expressive communication), motor (gross and fine), social-emotional and adaptive behavior. The first three are assessed through direct observation of the child in test situations, while the last two are assessed through questionnaires to be completed by the main caregiver. These last two scales are considered complementary and are less

used in clinical and research settings. Bayley-III motor scale assesses axial motor abilities like sitting, standing up and walking, as well as fine motor control abilities. Its cognition scale assesses the child's performance in several areas, such as, visualization, memory and attention, while the language scale assesses two major aspects of language, receptive and expressive communication skills, including a child's ability to recognize sounds and receptive vocabulary; the expressive communication subtest assesses preverbal communication, vocabulary use and morpho-syntactic development (Bayley, 2006). Bayley-III does not provide an overall total score, but separate raw and scaled scores for each domain as well as composite scores and percentile ranks for each scale. At the end of the process, the development of the child is classified as being on one of seven levels (extremely low, borderline, low average, average, high average, superior or very superior), based on the American population (Bayley, 2006). Bayley-III normative data were collected in the US in 2004 with 1,700 children aged 16 days to 43 months and 15 days. The reliability coefficients for Bayley-III subtests are .86 for fine motor, .87 for receptive communication and .91 for cognitive, expressive communication and gross motor (Bayley, 2006).

The *Peabody Developmental Motor Scale 2 (PDMS-2)* is composed of six subtests that measure interrelated abilities in early motor development: Reflexes, Stationary, Locomotion, Object Manipulation, Grasping, and Visual-Motor Integration. PDMS-2 results give a Total Motor Quotient, as well as, a Gross Motor Quotient and a Fine Motor Quotient. It was designed to assess the axial and appendicular motor ability of children up to 6 years of age, and was normed on 2,003 children residing in 46 states of the US and one Canadian province. The PDMS-2 has very good to excellent internal consistency ($r = .89 - .97$), test-retest reliability ($r = .89 - .96$), and interrater reliability ($r = .96 - .99$). Validity was examined for age differentiation. The correlation coefficients determined for 12-month age intervals ranged from $r = .80$ to $.93$, indicating that the subtests were associated with age, consistent with the developmental pattern of motor behaviors (Connolly, McClune, & Gatlin, 2012). At the time of the data collection, no instrument to assess motor development had been translated, adapted or validated for use in Brazil so the English version of PDMS-2 was used.

The Brazilian version of the *Visualization and Reasoning Battery of the Leiter International Performance Scale Revised - Leiter-R* is a nonverbal intelligence measurement tool that can be used in children starting at 2 years of age. It includes 6 subtests to assess visual processing and fluid reasoning of preschoolers: Figure-Ground (to evaluate visual discrimination and exploration), Form Completion (to assess visual synthesis ability), Matching, Classification (to evaluate the child's categorization capacity), Sequential Order (to assess sequential reasoning) and Repeated Patterns (to assess inductive reasoning). The translated version of this instrument has good validity and reliability for preschoolers. The Spearman-Brown coefficients ranged from .85 to .94 and Cronbach's alpha between .81 and .86 for the Leiter-R subtests, indicating good accuracy (Mecca, Antonio, Seabra, & Macedo, 2014). The Leiter-R predicted 24% of the arithmetic performance and

almost 31% of the read performance in schoolers (Mecca, Jana, Simões, & Macedo, 2015).

The *Language Development Survey (LDS)* checklist is a questionnaire, developed in Brazil, assessing expressive vocabulary by checking which words a child uses spontaneously. The mother/caregiver chose these from a list of 309 words categorized into 14 semantic groups, compiled from lexical development studies. This test is standardized for children aged 2 – 6 (Capovilla & Capovilla, 1997). The LDS manifested excellent concurrent validity with a brief direct screening measure of expressive vocabulary. The LDS test-retest reliability was from .97 to .99. The LDS correlated highly with The Reynell Receptive and Expressive Language Scale scores, The Bayley Mental Development Index and The Vineland Adaptive Behavior Composite (.66–.87). Sensitivity was > 80%, specificity was > 85%, positive predictive value, and negative predictive value between the LDS screening and the follow-up Reynell Expressive Language Scale were generally impressive (Rescorla & Alley, 2001).

The *Peabody Picture Vocabulary Test (PPVT)* consists of 144 items and evaluates the receptive vocabulary ability of children between 2 years and 6 months and 18 years of age. The PPVT was translated, adapted, validated and standardized to the Brazilian preschoolers (Capovilla & Capovilla, 1997). It covers a broad range of receptive vocabulary levels, from content areas (e.g., actions, vegetables, tools) and parts of speech (nouns, verbs, or attributes) across all levels of difficulty (Macedo, Capovilla, Duduchi, D'Antino, & Firmo, 2006). The test can be scored by hand or by computer. The internal consistency reliability is .94; the test-retest reliability is .93. The validity correlations with *EVT-2*: $r = .82$ (Dunn & Dunn, 1997).

Procedure

Data collection. All tests were performed individually, in the presence of a daycare teacher, at the place and time that were most convenient for the child. A professional trained in the Bayley-III scales conducted all the evaluations (except in the second phase of test-retest assessments), which took an average of 60 minutes per child. All other tests (PDMS-2, Leiter-R, LDS and PPVT) were performed by trained psychologists and lasted an average of 2 hours and 30 minutes per child. These four instruments were conducted and interpreted according to the age group of the child, using data from validation and normative studies. Out of the 207 participants, 81 were also tested with the PDMS-2, 58 with the Leiter-R, 69 with the LDS and PPVT language tests and 10 participated in the test-retest. Data collection took nine months to complete, from January to September 2012.

Data analysis. The raw scores of each of the Bayley-III scales and the total scores of the Leiter-R were used for descriptive and inferential analyses. Spearman correlation coefficients were calculated to assess convergent validity between the Bayley-III scales and the other instruments. Coefficients between .70 and 1 were considered to be of high magnitude; between .40 and .69 to be of moderate magnitude; and between .10 and .39 to be of low magnitude (Dancey & Reidy, 2013).

Exploratory factor analysis was used to assess the internal structure of the instrument. Principal component and oblique

rotation techniques were employed. This type of rotation is usually employed when there is a high correlation between subtests (Hair, Black, Babin, Anderson, & Tatham, 2009). For applicability, the following criteria were considered: Kaiser-Meyer-Olkin values $> .70$ significant Bartlett spherical test results ($p \leq .001$). *Eigen* values greater than or equal to one were used to select the number of components (Marôco, 2007).

In order to verify the adequacy of the factorial structure, a confirmatory factor analysis (CFA) was performed in accordance with the original validation study of Bayley-III conducted with an American sample. The CFA was done using AMOS IBM SPSS® version 20. In this way, the adjustment indices for the 1 factor (5 subtests on a general factor) and the 3 factor model (2 motor subtests on the 1st factor; 2 language subtests on the 2nd factor; and the cognition scale on the 3rd factor) were verified. The adequacy of the confirmatory indices were considered according to the following criteria: (1) Root Mean Square Error of Approximation (RMSEA) $< .05$ (Hair et al., 2009), (2) Comparative Fit Index (CFI) $> .95$ (Hu & Bentler, 1999), and (3) Tucker Lewis Index (TLI) ideally $> .90$ (Bentler & Bonett, 1980).

To assess the reliability of the Brazilian Bayley-III, we evaluated the stability of the instrument based on Spearman correlation analyses between the first and second tests. A non-parametric test was used due to the small number of participants in the retesting conducted nine months after the first test. Internal consistency was assessed using Cronbach's alpha coefficients and the Split-Half method with the Spearman-Brown formula being used. Analyses were performed using IBM SPSS® version 21.0 and p -value $< .05$ were considered statistically significant. To evaluate stability, Spearman correlation coefficients were calculated to assess test-retest reliability.

Ethical Considerations

The study protocol was approved by the Research Ethics Committee of the Universidade Presbiteriana Mackenzie (CAAE n. 0041.0.027.000-11) and authorized by the two daycare centers. Written informed consent was obtained from the legal guardians of all participating children.

Results

Translation of the scale from English to Brazilian Portuguese was done by a researcher fluent in both languages specialized in special education and is experienced in the use of the Bayley-III. This translated version was first submitted to a panel (P1) of two specialists in child development who independently provided practical and semantic suggestions to improve the text. These suggestions were sent to a second panel (P2) comprising two other specialists in child development, who analyzed and reviewed, their suggestions to produce a preliminary Brazilian version of the instrument. At this stage, some modifications were necessary to culturally adapt the Brazilian version of the scale, especially in respect of the traditional children's games and songs used in the language scales, to ensure the adequacy of the translated version. This preliminary Brazilian Bayley-III was then back-translated to English by another individual fluent in both languages. The

back-translated and the original English versions were sent to P2 who analyzed and made a few minor adjustments to create the final version of the instrument. This text was sent back to the authors of the original American scale who analyzed and approved the final official Brazilian version of the Bayley-III, specifically for use of the cognition, language and motor scales with children between 12 and 42 months of age.

The following section presents evidence of the validity and reliability of the results of the Brazilian version of Bayley-III. Table 1 presents mean and standard deviations, the minimum and maximum scores of participants in Bayley-III, PDMS-2, Leiter-R, LSD and PPVT.

Table 1
Descriptive Statistics of Measures

Instrument	<i>M (SD)</i>	Min	Max
Bayley-III			
Receptive Language	26.87 (10.09)	17	49
Expressive Language	27.58 (8.16)	20	47
General Language	54.45 (18.10)	37	94
Fine Motor	33.43 (9.20)	22	64
Gross Motor	48.49 (8.68)	35	71
Global Motor	81.93 (17.78)	57	135
Cognition	64.78 (10.85)	53	88
PDMS-2			
Prehension	12.60 (4.95)	5	24
Perceptual-motor integration	35.63 (13.32)	3	61
Static positioning	13.16 (2.71)	4	16
Locomotion	53.15 (22.73)	6	86
Object manipulation	15.11 (7.55)	0	28
General fine motor	48.23 (17.43)	9	83
General gross motor	81.42 (32.56)	10	130
PDMS-2 Total	129.65 (49.56)	19	212
Leiter-R_Raw Score	54.38 (17.92)	11	88
LDS	121.22 (53.40)	50	256
PPVT	27.86 (7.38)	13	44

Note. LDS: Language Development Survey Checklist. PPVT: Peabody Picture Vocabulary Test.

Spearman correlation tests were conducted between the raw scores in Bayley-III for the motor, cognition and language domains in relation to the other instruments which assess the same domains. The results showed that there was a significant and strong positive correlation between the Bayley-III fine, gross and global motor scales scores and the specific and general domain scores of the PDMS-2 (Table 2). Bayley-III cognition domain scores were overall positively correlated with the subtests and total Leiter-R scores. There was a moderate correlation between Bayley-III scores and the subtests Figure-Ground, Form Completion,

Matching and Classification. The first two subtests assess visual processing, including discrimination and synthesis, and the last two assess the ability to categorize color, shapes, sizes or semantic associations. There was a low correlation between the Bayley-III cognition scores and the Leiter-R Sequential Order subtest and no significant correlation with the subtest Repeated Patterns which assesses inductive reasoning. This result suggests that the Bayley-III cognition

scale is more related with the performance of categorization and visualization tasks than with sequential or inductive reasoning (Table 2). As also presented in Table 2, there was a strong correlation between Bayley-III receptive, expressive and general language scores and LDS and PPVT scores. The high degree of correlation between the different Bayley-III domain scores and the various other instruments indicates convergent validity.

Table 2
Correlation Analyses Between Bayley-III Cognition and Language Scores With Leiter-R, LDS and PPVT Scores

	Bayley-III						
	Cognition Scale	Receptive Language	Expressive Language	General Language	Fine motor	Gross Motor	Global motor
Leiter-R							
Figure-Ground	.56*						
Form Completion	.48*						
Matching	.60*						
Sequential Order	.37*						
Repeated Pattern	.14						
Classification	.60*						
Total Leiter-R score	.61*						
Language							
LDS		.94*	.96*	.96*			
PPVT		.86*	.85*	.86*			
PDMS-2							
Prehension					.84**	.84**	.84**
Perceptual-motor Integration					.83**	.89**	.86**
Static positioning					.71**	.77**	.74**
Locomotion					.84**	.89**	.87**
Object manipulation					.88**	.93**	.91**
General fine motor					.87**	.92**	.90**
General gross motor					.85**	.90**	.88**
PDMS-2 Total					.87**	.92**	.90**

* $p \leq .05$. ** $p \leq .01$.

There was also a strong positive correlation among specific scales (score domains) of the Bayley-III tool. The strongest correlations were between the receptive and

expressive language domains and the gross and fine motor domains (Table 3). The fine motor domain had the strongest correlation with the cognition domain.

Table 3
Correlation Analyses Between Individual Bayley-III Score Domains

	Receptive L.	Expressive L.	General L.	Fine motor	Gross motor	General motor	Cognition
Receptive L.		.96*	.99*	.89*	.82*	.86*	.71*
Expressive L.			.89*	.88*	.82*	.86*	.77*
General L.				.89*	.83*	.87*	.75*
Fine motor					.97*	.99*	.83*
Gross motor						.99*	.77*
General motor							.81*

Note. L. = language.

* $p \leq .05$.

The criteria for the factor analysis were met with KMO values = .764 and Bartlett's Sphericity > .001. The exploratory factor analysis used the component and oblique rotation techniques (direct oblmin) and identified only one component with an *eigenvalue* of 4.29, which explained 86% of the variance. This indicates that the instrument in fact assesses a general dimension of child development. These five components had a high factorial weight, loading a single factor: Fine motor = .96, Gross motor = .91, Receptive language = .91, Expressive language = .94 and Cognition = .88.

The results obtained from the CFA showed high factor loadings for each scale in general factor, considering the model with one factor. All correlations were significant ($p \leq .001$), as illustrated in Figure 1. The good-fit index for the model indicated a factor with RMSEA < .001; CFI = 1.00; TLI = 7.73. These results show good fit index for the model with just one factor. It was not possible to estimate the 3 factors model with the sample data of the present study.

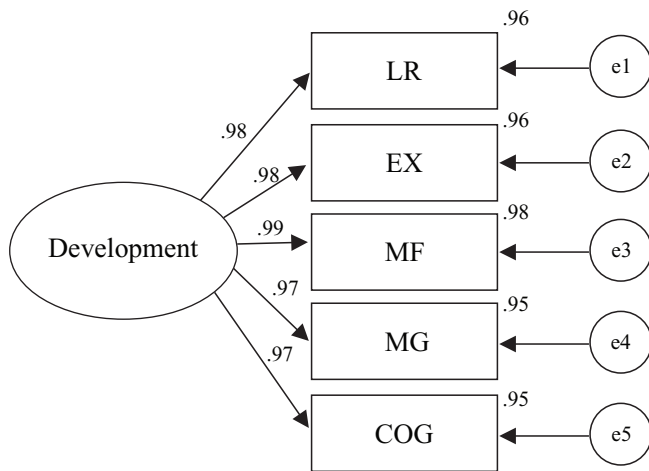


Figure 1. Confirmatory Factory Analysis according to a model of one factor (5 subtests on a general factor). Note. LR = receptive language; EX = expressive language; MF = fine motor; MG = gross motor; COG = cognition.

Reliability of the Bayley-III tool was assessed by measuring the stability (test-retest) of all domains. It was not possible to assess the test-retest scores of the expressive and receptive language domains because the children had the same score in the first assessment and therefore we could not calculate a variance. We did not find a significant positive correlation between the test-retest scores for cognition ($Rho = -.34; p = .449$) or for gross motor ($Rho = -.39; p = .375$). There was a positive correlation for fine motor scores between the two assessments ($Rho = .89; p = .007$).

Table 4 presents the internal consistency results for each Bayley-III domain and also for the total score using Cronbach alpha coefficients and the Split-Half method using the Spearman-Brown formula. The results indicate low measurement errors for the Bayley subscales and the tool in general.

Table 4
Internal Consistency Analyses of the Brazilian Version of Bayley-III Scales

Variable	Cronbach's alpha coefficient	Split-Half by Spearman-Brown	Correlation between the two halves
Fine motor	.95	.98	.97
Gross motor	.95	.99	.98
General motor	.98	-	-
Receptive language	.96	.99	.99
Expressive language	.96	.98	.97
General language	.97	-	-
Cognition	.96	.98	.96
Bayley-III	.90	-	-

Discussion

The increasing number of recent Brazilian studies that used the Bayley-III scales indicates the importance and usefulness of this instrument in the diagnosis of motor, cognitive and language delays in young Brazilian children (Ferreira et al., 2014; Hentges et al., 2014). However, the authors of previous studies used the original English version of the Bayley-III or non-validated translations that did not follow the guidelines for the process of cross-cultural adaptation (Hambleton & Patsula, 1999) and with unknown psychometric properties. These limitations could have influenced the reliability of the scores and the interpretation of the results provided in these studies (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). Taking this into account, one of the main goals of the present study was to translate and adapt the Bayley-III to Portuguese following the best evidence-based guidelines for the translation, cross-cultural adaptation and assessment of psychometric properties, in addition to using the equivalence criteria proposed by Herdman et al. (1998). Future studies using the translated and adapted version of Bayley-III can help to improve it further and contribute to the research and clinical fields, helping health professionals to better identify young children at risk of developmental delay.

We first assessed the convergent validity of the Bayley-III scales by correlating it with other instruments that, in theory, measure the same abilities as in its domains (AERA et al., 1999). We found a high positive correlation between the Brazilian version of the Bayley-III motor domains (total, fine and gross) and the specific and general PDMS-2 scores. Thus, both instruments seem to be of good quality, but Bayley-III is briefer and easier to administer.

Similar results were identified in the language domains of the Bayley-III tool compared with LDS and PPVT scores. These results indicate that Bayley-III evaluates in a different way, almost the same skills as the other instruments. It is important to note that LDS is an indirect assessment

tool applied to a child's caregiver. But the PPVT is an instrument composed of items in the same format, i.e., the child's receptive vocabulary is assessed from the figures of choice when they hear the target stimulus. One of Bayley's advantages is not only that it makes a direct assessment of the child and looks at younger age groups in comparison with others instruments, it also assesses a range of diverse items, including the reaction to ambient sounds, recognition of familiar words and more complex levels like sentence comprehension.

There was a moderate positive correlation between the Brazilian Bayley-III cognition scores and the Leiter-R scores. Since the Leiter-R tool has subtests that assess different cognitive abilities (Mecca et al., 2014) the correlation differed between specific subtests. Stronger correlations were found for categorization and visual processing abilities than for tasks related to fluid intelligence. The common variance observed between the Bayley-III Cognitive Scale and the visual processing and categorization tasks of the Leiter-R was expected, since the items in Bayley-III require the child to have the capacity to visual explore stimuli and knows how to sort them according to certain categories. These skills are developed very early on and increase significantly during the preschool years (Mecca et al., 2014). On the other hand, there are few items in the Bayley-III requiring sequential and inductive reasoning, which are the last items in the Cognitive Scale and therefore the most difficult, because they are the skills that develop more fully from 5 to 6 years of age (Mecca et al., 2014).

The exploratory factor analyses of the internal structure of the Brazilian version of the Bayley III scales found that a single component explained 86% of the variance and this result was corroborated by good fit indices shown by the CFA. This result allows us to conclude that the total score of this version of the Bayley-III reflects the general component of child development and that the total score of this instrument can be interpreted as a global measure of child development.

Due to the high correlation between specific domains, future studies are needed to confirm if these specific factors are present in other samples and ages. If future studies corroborate our findings it may be possible to produce a reduced version of the Bayley-III, decreasing the number of items per domain or even excluding entire domains. This would be an advantage in a version adapted to Portuguese, since several studies show the benefit of using brief or short assessment tools (Coutinho & Nascimento, 2010; Mello et al., 2011). A short version of the Bayley-III could reduce the time required not only for research but also in the case of a need for screening when there is suspected developmental delay.

In the present study, the model with three factors cannot be estimated and the one factor model fits better than reported in Bayley (2006). These findings differ from those reported by Bayley, who identified three different factors for language, cognitive and motor performance. This discrepancy can in part be explained due to the much larger number of participants in Bayley's study and to the type of analysis performed. However, in the original Bayley-III manual, it remains unclear whether the raw or standardized

scores were used in the analysis, and whether they were based on the individual items of the instrument or on the total scores. In addition, there may also be an issue in terms of the differences between the studies related to the selection of participants. The original study used a stratified sample of 1,700 children (Bayley, 2006), whereas in the present study we used a convenience sample whose participants had similar socioeconomic aspects. This is one of the main limitations of the study, especially regarding the generalization and comparison of findings. An additional limitation of the current study is that the children's health information (chronic diseases and developmental disorders), used for exclusion criteria were based on the records of the daycare centers without any independent clinical evaluation. Besides, we did not performed analyses based on sociodemographic characteristics of the children or their families, because the group was considered mostly homogenous. We also did not collect data about environmental stimulation, although it is important to note that all the participants were exposed to the same level of stimulation in the day care centers, since all of them stayed there full time.

The reliability of the Brazilian version of the Bayley III was good, with excellent internal consistency and item homogeneity (AERA et al., 1999). The results of the score stability were less robust. These data may be due to the age group of our participants. In very young children, the development of abilities is not as stable as in older preschoolers and in school age children (Griffiths, 1996). The lack of stability of scores over time indicates that Bayley-III may not be a good tool for identifying the effects of interventions or for predicting future performance with the same scale.

The Brazilian version of the Bayley-III instrument had high convergent validity and good internal consistency and item homogeneity for children aged 12-42. This version can be useful for research purposes. Further studies with this version of the Bayley-III are needed, involving larger random samples from different regions of the country, as well as cohort studies to establish development curves comparing the performance in different age groups. There is also a need to perform more studies to assess the internal structure of this version of the Bayley-III using item analyses instead of total scores, as well as confirmatory factor analyses according to age groups with a higher number of participants, as was done in the original version of this instrument.

Finally, this first study on the psychometric properties of the Brazilian version of the Bayley-III instrument will be useful for future studies comparing the development of normal versus high-risk children or those with specific clinical conditions. Thus, the present study contributes to advances in the assessment of child development in Brazil, a country without any similar validated tools.

References

- Acton, B. V., Biggs, W. S., Creighton, D. E., Penner, K. A., Switzer, H. N., Thomas, J. H. P., . . . Robertson, C. M. (2011). Overestimating neurodevelopment using

- the Bayley-III after early complex cardiac surgery. *Pediatrics*, 128(4), e794-e800. doi:10.1542/peds.2011-0331
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, TX: Pearson.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588-606. doi:10.1037/0033-2909.88.3.588
- Capovilla, F. C., & Capovilla, A. G. S. (1997). Desenvolvimento linguístico da criança dos dois aos seis anos: Tradução e standardização do Peabody Picture Vocabulary Test de Dunn & Dunn, e da Language Development Survey de Rescorla [Child's language development from two to six years: Translation and standardization of the Dunn & Dunn Peabody Picture Vocabulary Test, and Rescorla's Language Development Survey]. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, 1(1), 353-380.
- Connolly, B. H., McClune, N. O., & Gatlin, R. (2012). Concurrent validity of the Bayley-III and the Peabody Developmental Motor Scale-2. *Pediatric Physical Therapy*, 24(4), 345-352. doi:10.1097/PEP.0b013e318267c5cf
- Coutinho, A. C. A. D. M., & Nascimento, E. (2010). Formas abreviadas do WAIS-III para avaliação da inteligência [Abbreviated forms of the Wais-III for intelligence assessment]. *Avaliação Psicológica*, 9(1), 25-33.
- Dancey, C. P., & Reidy, J. (2013). *Estatística sem matemática para psicologia [Statistics without Maths for Psychology]* (L. Viali, Trans., 5th ed.). Porto Alegre, RS: Penso.
- Dunn, L. M., & Dunn, L. M. (1997). *PPVT: III-A Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.
- Eickmann, S. H., Malkes, N. F. A., & Lima, M. C. (2012). Psychomotor development of preterm infants aged 6 to 12 months. *São Paulo Medical Journal*, 130(5), 299-306. doi:10.1590/S1516-31802012000500006
- Fernandes, L. V., Goulart, A. L., Santos, A. M. N., Barros, M. C. M., Guerra, C. C., & Kopelman, B. I. (2012). Neurodevelopmental assessment of very low birth weight preterm infants at corrected age of 18-24 months by Bayley III scales. *Jornal de Pediatria*, 88(6), 471-478. doi:10.2223/JPED.2230
- Ferreira, R. C., Mello, R. R., & Silva, K. S. (2014). Neonatal sepsis as a risk factor for neurodevelopmental changes in preterm infants with very low birth weight. *Jornal de Pediatria*, 90(3), 293-299. doi:10.1016/j.jpmed.2013.09.006
- Fleuren, K. M. W., Smit, L. S., Stijnen, T. H., & Hartman, A. (2007). New reference values for the Alberta Infant Motor Scale need to be established. *Acta Paediatrica*, 96(3), 424-427. doi:10.1111/j.1651-2227.2007.00111.x
- Griffiths, R. (1996). *Griffiths mental development scales - Revised*. High Wycombe, UK: The Test Agency.
- Guedes, D. Z., Primi, R., & Kopelman, B. I. (2011). BINS validation - Bayley neurodevelopmental screener in Brazilian preterm children under risk conditions. *Infant Behavior & Development*, 34(1), 126-135. doi:10.1016/j.infbeh.2010.11.001
- Hair, J. F., Jr., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados [Multivariate data analysis]* (A. S. Sant'Anna, Trans., 6th ed.). Porto Alegre, RS: Bookman.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1(1), 1-13.
- Hentges, C. R., Silveira, R. C., Procianny, R. S., Carvalho, C. G., Filipouski, G. R., Fuentesfria, R. N., . . . Terrazan, A. C. (2014). Association of late-onset neonatal sepsis with late neurodevelopment in the first two years of life of preterm infants with very low birth weight. *Jornal de Pediatria*, 90(1), 50-57. doi:10.1016/j.jpmed.2013.10.002
- Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. *Quality of Life Research*, 7(4), 323-335. doi:10.1023/A:1024985930536
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi:10.1080/10705519909540118
- Macedo, E. C., Capovilla, F. C., Duduchi, M., D'Antino, M. E. F., & Firmo, L. S. (2006). Avaliando linguagem receptiva via teste de vocabulário por imagens peabody: Versão tradicional versus computadorizada [Evaluating receptive language by peabody picture vocabulary test: Computerized versus traditional versions]. *Psicologia: Teoria e Prática*, 8(2), 40-50.
- Marôco, J. (2007). *Análise estatística: Com utilização do SPSS [Statistical analysis: Using SPSS]*. Lisboa, Portugal: ReportNumber.
- Mecca, T. P., Antonio, D. A. M., Seabra, A. G., & Macedo, E. C. (2014). Parâmetros psicométricos da Escala Internacional de Inteligência Leiter-R para crianças pré-escolares [Psychometric parameters of the Leiter-R International Performance Scale for preschool children]. *Avaliação Psicológica*, 13(1), 125-132.
- Mecca, T. P., Jana, T. A., Simões, M. R., & Macedo, E. C. (2015). Relação entre habilidades cognitivas não-verbais e variáveis presentes no contexto educacional [Relationship between non-verbal cognitive abilities and variables in the educational context]. *Psicologia Escolar e Educacional*, 19(2), 329-340. doi:10.1590/2175-3539/2015/0192844
- Mello, C. B., Argollo, N., Shayer, B. P. M., Abreu, N., Godinho, K., Durán, P., . . . Bueno, O. F. A. (2011).

Versão abreviada do WISC-III: Correlação entre QI estimado e QI total em crianças brasileiras [Abbreviated version of the WISC-III: Correlation between estimated IQ and global IQ of Brazilian children]. *Psicologia: Teoria e Pesquisa*, 27(2), 149-155. doi:10.1590/S0102-37722011000200002

- Milne, S., McDonald, J., & Comino, E. J. (2012). The use of the Bayley scales of infant and toddler development III with clinical populations: A preliminary exploration. *Physical & Occupational Therapy in Pediatrics*, 32(1), 24-33. doi:10.3109/01942638.2011.592572
- Moore, T., Johnson, S., Haider, S., Hennessy, E., & Marlow, N. (2012). Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *The Journal of Pediatrics*, 160(4), 553-558. doi:10.1016/j.jpeds.2011.09.047
- Pinto, E. B., Vilanova, L. C. P., & Vieira, R. M. (1997). *O desenvolvimento do comportamento da criança no primeiro ano de vida: Padronização de uma escala para a avaliação e o acompanhamento [Children's behavior development in the first year of life: Standardization of a scale to evaluate and monitor]*. São Paulo, SP: Casa do Psicólogo.
- Rescorla, L., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research*, 44(2), 434-445. doi:10.1044/1092-4388(2001/035)
- Reuner, G., Fields, A. C., Wittke, A., Löpprich, M., & Pietz, J. (2013). Comparison of the developmental tests Bayley-III and Bayley-II in 7-month-old infants born preterm. *European Journal of Pediatrics*, 172(3), 393-400. doi:10.1007/s00431-012-1902-6
- Santos, R. S., Araújo, A. P., & Porto, M. A. S. (2008). Early diagnosis of abnormal development of preterm newborns: Assessment instruments. *Jornal de Pediatria*, 84(4), 289-299. doi:10.1590/S0021-75572008000400003
- Silveira, K. A., & Enumo, S. R. F. (2012). Riscos biopsicossociais para o desenvolvimento de crianças prematuras e com baixo peso [Biopsychosocial risks to development in preterm children with low birth weight] *Paidéia (Ribeirão Preto)*, 22(53), 335-345. doi:10.1590/S0103-863X2012000300005
- Yu, Y. T., Hsieh, W. S., Hsu, C. H., Chen, L. C., Lee, W. T., Chiu, N. C., & Jeng, S. F. (2013). A psychometric study of the Bayley Scales of Infant and Toddler Development-3rd Edition for term and preterm Taiwanese infants. *Research in Developmental Disabilities*, 34(11), 3875-3883. doi:10.1016/j.ridd.2013.07.006

Vanessa Madaschi holds a Master's degree in Developmental Disorders from Universidade Presbiteriana Mackenzie.

Tatiana Pontrelli Mecca is an Associate Professor of the Centro Universitário FIEO.

Elizeu Coutinho de Macedo is an Associate Professor of the Universidade Presbiteriana Mackenzie.

Cristiane Silvestre de Paula is an Associate Professor of the Universidade Presbiteriana Mackenzie.

Received: Mar. 20, 2015

1st Revision: Nov. 6, 2015

Approved: Nov. 9, 2015

How to cite this article:

Madaschi, V., Mecca, T. P., Macedo, E. C., & Paula, C. S. (2016). Bayley-III scales of infant and toddler development: Transcultural adaptation and psychometric properties. *Paidéia (Ribeirão Preto)*, 26(64), 189-197. doi:10.1590/1982-43272664201606