



Application of neurodevelopmental screening to a sample of South American infants: The Bayley Infant Neurodevelopmental Screener (BINS)[☆]

Ann Marie McCarthy^{a,*}, George L. Wehby^b, Sheila Barron^c, Glen P. Aylward^d, Eduardo E. Castilla^{e,f}, Lorette C. Javois^g, Norman Goco^h, Jeffrey C. Murrayⁱ

^a College of Nursing, The University of Iowa, 50 Newton Road, CNB 344, Iowa City, IA 52242, United States

^b College of Public Health, E204 GH, The University of Iowa, Iowa City, IA, United States

^c Statistics Outreach Center, College of Education, The University of Iowa, Iowa City, IA, United States

^d Division of Developmental and Behavioral Pediatrics/Psychology, Southern Illinois University School of Medicine, Springfield, IL, United States

^e Instituto Nacional de Genética Médica Populacional (INAGEMP) and Estudio Colaborativo Latino Americano de Malformaciones Congénitas (ECLAMC) at CEMIC: Centro de Educación Médica e Investigación Clínica, Buenos Aires, Argentina

^f Laboratório de Epidemiologia de Malformações Congênitas, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

^g Developmental Biology, Genetics and Teratology Branch, Center for Developmental Biology and Perinatal Medicine, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, United States

^h RTI International, Research Triangle Park, NC, United States

ⁱ Department of Pediatrics, The University of Iowa, Iowa City, IA, United States

ARTICLE INFO

Article history:

Received 14 March 2011

Received in revised form

29 September 2011

Accepted 12 December 2011

Keywords:

Developmental screening

Infants

South America

Bayley Infant Neurodevelopmental Screener

ABSTRACT

Objective: To evaluate the utility of the Bayley Infant Neurodevelopmental Screener (BINS), standardized in the US, for South American infants, 3–24 months of age.

Methods: Thirty-five physicians administered the BINS to 2471 South American infants recruited during routine well-child visits, 578 (23%) from Brazil and 1893 (77%) from six other South American countries. The BINS was translated into Spanish and Portuguese and participating physicians were trained to administer the BINS. Physician inter-rater agreement with training tapes was 84.4%; test–retest reliability for age item sets ranged from 0.80 to 0.93 (Pearson's *r*). Infants were classified into being at low, moderate, or high risk for developmental delay or neurological impairment based on their total BINS score. The sample was stratified by infant's age, sex and language (Spanish and Portuguese). The BINS scores were compared to the scores of the US infant sample used to standardize the BINS. **Results:** Female infants performed higher than male at 16–20 months and 21–24 months; male infant scores were more variable at 5–6 months. Scores on only two items were significantly different between Spanish and Portuguese speaking participants. South American scores were typically significantly higher than the US sample, and a lower proportion of infants were classified as being at high risk in the South American sample than in the US standardization sample.

Conclusion: Overall, the results of this study indicate that the BINS is feasible and appropriate for neurodevelopmental screening in South America. Further studies are needed to confirm the BINS utility in South America, including its use with a clinical sample.

© 2012 Elsevier Inc. All rights reserved.

[☆] Funding: NIH/NICHD grant 1U01 HD-40561-05S2.

* Corresponding author. Tel.: +1 319 335 7087; fax: +1 319 335 5326.

E-mail address: ann-mccarthy@uiowa.edu (A.M. McCarthy).

1. Introduction

Developmental disabilities are common disorders that impose significant physical, mental and learning limitations on affected children. In the United States (US), about 13% of children 3–17 years of age have at least one developmental disability and about 1.6% of children have three or more developmental disabilities (Boulet, Boyle, & Schieve, 2009). In the US, children under three years of age with developmental disabilities are eligible to receive early intervention services through programs such as the Individuals with Disabilities Education Act (IDEA) (*Individuals with Disabilities Education Improvement Act, 2004*).

Identifying infants at risk for developmental disabilities is the first step in providing them with services to maximize their physical and cognitive abilities and to minimize complications. The American Academy of Pediatrics recommends that pediatricians screen all infants and children during routine office visits for developmental problems (*Council on Children with Disabilities, 2006*). In the US, the emphasis has shifted to screening for disabilities at a younger age, birth to 2 years, with the passage of the IDEA amendments (*Individuals with Disabilities Education Act, 1997, 2004*). Certain health conditions such as low birth weight, preterm birth, perinatal infection and birth defects increase the risk for developmental difficulties. For example, children born with birth defects are almost 27 times more likely to have a developmental disability by age 7 compared to children who were not born with a birth defect (Jelliffe-Pawlowski, Shaw, Nelson, & Harris, 2003).

Robust estimates of the prevalence of development disabilities in less developed countries are rare. Rates are expected to be at least similar to if not higher than those in developed countries given the overall higher prevalence of most diseases of early childhood in less developed countries compared to developed countries. The availability of adequate screening for developmental disabilities is limited in less developed countries where expenses on health are significantly lower than developed countries. For example, total health expenditures per capita in Latin America are less than one-tenth of those in developed nations (Govindaraj, Chellaraj, & Murray, 1997). Improved economic status has positive effects on child development in both developed and developing countries (Currie, 2009; Paxson & Schady, 2007) and may attenuate the negative effects of early developmental problems on future outcomes (Feinstein, 2003). Therefore, developmental disabilities may have larger adverse effects on future health and socioeconomic outcomes in less developed countries due to the higher poverty rates.

One factor that limits screening in less developed countries is that, unlike the US where several standardized instruments are available to screen and assess infant development, few instruments have been standardized or evaluated for use in less developed countries. Instruments that are translated for use with ethnic minority groups in the US sometimes are used in other countries without proper evaluation. This limits the ability of health care practitioners in less developed countries to carry out systematic screening procedures and to refer children for available remediation programs even when the costs of such screening are minimal. An important question is whether instruments standardized in the US are useful and applicable in less developed countries. Assessing the utility of these instruments for use in less developed settings is essential to enhance the capacity of health professionals to screen for developmental disabilities. Primary care providers play a crucial role in the identification of children with developmental difficulties through frequent monitoring of development, identifying at-risk children who are outside the normal range of development, and referring them for further developmental assessment and treatment as needed. Increasing the availability of developmental screening instruments in less developed countries is also needed for quantifying at the population-level the developmental burdens of several early life health deficits such as low birth weight (Boulet, Schieve, & Boyle, 2009), preterm birth (Anderson & Doyle, 2008), infections, birth defects and others that may place a large burden on child development. In South America, there is a real need for standardized developmental screening instruments that can be used by primary care providers. Identifying the utility of existing instruments that are used in the US to screen infants for risk for developmental problems is needed to expand the capacities of pediatricians and primary care physicians to screen for neurodevelopmental problems in South America.

The Bayley Infant Neurodevelopmental Screener (BINS) is a developmental screening measure that offers an alternative to detailed assessment for infants 3–24 months of age. It can be administered by a wide range of health professionals with limited training and in an acceptable time frame for screening. In the United States, the BINS is cited as a useful screening instrument by the American Academy of Pediatrics (*Council on Children with Disabilities, 2006*), and others (Dobrez et al., 2001). It has been employed as a standard tool in developmental follow-up clinics (Aylward & Verhulst, 2000; Leonard, Piecuch, & Cooper, 2001; Macias et al., 1998) as well as in general pediatric populations (Blackman, 1999; Dobrez et al., 2001). The BINS has been found to be particularly valuable in screening high risk infants for developmental delay (Aylward & Verhulst, 2000; Constantinou, Adamson-Macedo, Mirmiran, Ariango, & Fleisher, 2005; Gucuyener et al., 2006).

The purpose of this project was to evaluate the utility of the Bayley Infant Neurodevelopmental Screener (BINS), an instrument standardized in the US, for South American infants, 3–24 months of age. The specific objectives of this study were to:

1. Assess the distribution of South American BINS scores by age, sex, and language, and
2. Compare South American distributions of BINS scores and developmental risk categories with US norms.

Table 1
Number and percent of participants from each country.

Country	N	Percent
Argentina	754	30.51
Brazil	578	23.39
Ecuador	490	19.83
Chile	404	16.35
Bolivia	130	5.26
Colombia	66	2.67
Venezuela	49	1.98
Total	2471	99.99

2. Materials and methods

2.1. Setting

An existing physician network affiliated with the Latin American Collaborative Study of Congenital Malformations (Estudio Colaborativo Latino Americano de Malformaciones Congenitas, ECLAMC) was used to recruit the study sample and obtain data. Since 1967, ECLAMC has provided surveillance of birth defects throughout South America (Castilla & Orioli, 2004) and a unique research structure for numerous child health studies (López-Camelo, Campaña, Santos, & Poletta, 2006; Wehby et al., 2006; Wehby, Murray, Castilla, Lopez-Camelo, & Ohsfeldt, 2009; Wehby, Castilla, & Lopez-Camelo, 2010). The physicians voluntarily report to ECLAMC the occurrence of birth defects among infants born in affiliated hospitals. These physicians also routinely care for infants and children who are healthy as well as infants with a range of health conditions in their general pediatric practices.

2.2. Participants

Children in this study were recruited in 2005 and 2006 from the practices of the participating ECLAMC-affiliated physicians. Inclusion criteria were being a healthy child between 3 and 24 months of age and being seen by the physician for routine well-child care. Exclusion criteria were maternal hospitalization for complications when pregnant with the child, low or unknown birth weight, preterm birth, 5-min Apgar score < 6, twin (or multiple birth), oxygen given at birth, spending more than 5 days in the hospital after birth, admission to neonatal intensive care unit, a chronic illness that requires regular treatment and/or medicine for more than 2 weeks (excluding otitis media and allergies), documented prior developmental delay, and major surgery. The mother or a primary caregiver of the child had to be with the child at the pediatric visit.

The study sample included 2471 infants in South American countries recruited by 35 physicians, 578 (23%) from Portuguese-speaking Brazil (recruited by 8 physicians) and 1893 (77%) from six Spanish-speaking countries (recruited by 27 physicians). The participating countries and number of participants per country are found in Table 1. The number of participants per the six BINS age groups (3–4 months, 5–6 months, 7–10 months, 11–15 months, 16–20 months, 21–24 months which are described below) ranged from 377 to 487. The study sample was stratified by language, sex and age group and was projected based on an anticipated enrollment capacity of about 60 children per pediatrician over the course of the study (10 children, 5 males and 5 females, in each of the six BINS age groups). For comparison, the US sample used in standardizing the BINS included 600 infants across all six age groups between 3 and 24 months of age and 300 clinical cases (Aylward, 1995).

The study sites were located in geographically and socioeconomically diverse communities. The sample was roughly evenly divided between male ($n = 1241$, 50.2%) and female infants ($n = 1229$, 49.8%) (one child was missing data on sex). Almost half (48.0%) of the sample were first-borns, 28.4% had one sibling, 13.0% had two siblings, and 4.5% had more than two siblings. The self-reported ethnic ancestry of the child was African for 12.4% of the sample and Native for 44.9% of the sample (versus other which included primarily Latin-European Ancestry). The average age of the mother was 27.21 years (6.57 SD) with a range of 14–49. More than one-third of the mothers (35.4%) were in stable relationships and 13.7% identified themselves as single. For maternal education, 24.6% attended only a primary school, 50.0% had a secondary school education (partial or complete), and 25.4% had partial or complete university education. More than a third (36.5%) of the mothers reported being employed. This diversity provides inference that the sample is representative of large percentages of children in the study countries. Unfortunately, national-level data are not readily available on these characteristics for all study countries to compare to the sample characteristics. Nonetheless, the study employs a unique existing network of providers for selecting a large multi-country and geographically, socioeconomically, and demographically diverse sample of children.

2.3. Study measures

2.3.1. Bayley Infant Neurodevelopmental Screener (BINS)

The BINS (Aylward, 1995) was designed to identify infants, 3–24 months of age, at risk for developmental delays or neurological impairments by assessing four conceptual areas of ability: (a) basic neurological functions/intactness (posture, muscle tone, movement, asymmetries, abnormal indicators), (b) expressive functions (gross motor, fine motor, oral motor/verbal), (c) receptive functions (visual, auditory, verbal), and (d) cognitive processes (object permanence, goal-directedness, problem solving). The BINS is derived from the Early Neuropsychologic Optimality Rating Scales (ENORS) (Aylward, 1994) and also includes some items from the Bayley Scale of Infant Development, 2nd Edition (Bayley, 1993) as well as items that assess muscle tone and movement. This screening test identifies infants in need of further diagnostic evaluation and, with repeated administration, may be used to monitor the course of an infant's development (Aylward, 1995, 2010; Benish, 1998).

The BINS consists of six item sets grouped by age (3–4 months, 5–6 months, 7–10 months, 11–15 months, 16–20 months, 21–24 months), each containing 11–13 items. Each item in the BINS is scored "optimal" or "nonoptimal," based on a priori decision rules; the number of optimal responses for a given item set are then added to provide a summary score. The infant's total score is then compared to the norms in order to classify the infant into low, moderate, or high risk for developmental delay or neurological impairment (Aylward, 1995; Benish, 1998).

The BINS takes approximately 10 min to administer. The BINS is reported to have good internal consistency (0.73–0.85), test–retest reliability (Pearson's r ; 0.71–0.84), and inter-rater reliability (Cronbach's α ; 0.79–0.96) (Aylward, 1995). A concurrent validity for the BINS and Bayley Scales of Infant Development II (BSID-II) yielded Pearson correlation coefficients (r) that ranged from 0.43 to 0.82 for the Mental Developmental Index and 0.39–0.58 for the Psychomotor Developmental Index (Aylward, 2010). Studies of predictive validity showed that an early classification of an infant in the high-risk category was a strong indicator of future developmental problems (Aylward & Verhulst, 2008; Aylward, 1995, 1998).

2.3.2. Screening form and demographic data

A screening form that included questions on the exclusion criteria listed above was developed to screen for the appropriateness of a child to participate in this study. Once enrolled, a Demographic Data Form was used to collect information on the demographic and economic characteristics of the household and the child's birth and health history. A subgroup of participating physicians reviewed and approved the cultural appropriateness of these forms.

2.4. Procedure

The BINS scoring form and the relevant sections of the training manual were translated into Spanish and Portuguese. The translations were then reviewed for validity and accuracy. Next, all the physicians enrolling children in this study were trained in the administration of the BINS at a group meeting in Brazil by one of the study investigators (AMM). Participating physicians were provided BINS testing kits, translated study procedures and scoring forms. In addition, the reliability of the majority of physicians in screening a child's development was assessed before beginning recruitment by comparing their ratings of a set of videotaped case evaluations to the scores of one of the authors who developed and has significant expertise in using the BINS (GA). The physicians' agreement with the gold standard score was 84.4%.

Letters explaining the project, along with a verbal explanation from a staff member, were given to the parent or primary caregiver of the eligible children. After screening for eligibility based on the criteria described above, the study staff administered the informed consent document, in the appropriate language. The study physician evaluated the child's development with the BINS. In addition, for test–retest reliability, about 10% of the children from each age group were randomly selected to be retested no less than one week and no more than two weeks following the initial test. A total of 236 infants were retested at a mean number of 10.3 days between the first and second test. The test–retest reliability measured using Pearson's Product-Moment correlation coefficient (r) ranged from 0.8 for age group 3 to 4 months to 0.93 for age-group 21–24 months.

2.5. Data management

Eligibility and demographic/socioeconomic data were transmitted from the sites to the study's Data Center at RTI International (North Carolina). The study physicians used personal digital assistants (PDAs) to collect data on eligibility and demographic/socioeconomic characteristics and to keep an updated report on the numbers of children expected and enrolled by age group and sex. A program for random selection of enrolled children for test–retest reliability was also added to the PDAs. The PDA data collection system had built-in quality control checks including range and skip patterns. Data were evaluated for errors and inconsistencies, which were corrected through communication between the Data Center and the study pediatricians. The BINS data were collected on paper forms, which were double keyed and compared to identify and resolve data entry errors.

The youngest age inclusion criterion for our study was 3 completed months. However, the youngest age inclusion criterion for the US BINS standardization sample was 2 months and 15 days. Therefore, statistical adjustment was needed when comparing the 3 month group in our study to the 3 month group in the US standardization sample since no data were available in our study for ½ of the month (2.50–3.00). An adjusted BINS score distribution for the 3 month group was

Table 2
Means and SDs of South American BINS scores by age.^a

Number of BINS items	Age in months	N	Mean	Std. dev.	Minimum	Maximum
3–4 month form	11	152	9.29	1.08	5.00	11.00
	4	237	9.73	1.11	6.00	11.00
5–6 month form	13	157	10.40	1.74	4.00	13.00
	6	225	11.18	1.30	7.00	13.00
7–10 month form	13	97	9.99	1.87	4.00	13.00
	7	125	10.44	1.71	4.00	13.00
	8	154	11.42	1.21	7.00	13.00
	9	111	11.66	1.30	8.00	13.00
11–15 month form	11	61	8.77	1.37	6.00	11.00
	12	123	9.31	1.21	5.00	11.00
	13	92	9.52	1.35	5.00	11.00
	14	84	9.99	1.14	5.00	11.00
	15	90	10.39	0.99	6.00	11.00
16–20 month form	11	50	7.46	1.72	3.00	11.00
	16	62	8.44	1.37	5.00	11.00
	17	93	8.48	1.54	5.00	11.00
	18	92	8.90	1.62	4.00	11.00
	19	80	9.18	1.57	3.00	11.00
21–24 month form	13	55	10.15	2.12	3.00	13.00
	21	101	9.90	2.12	5.00	13.00
	22	128	10.77	2.03	5.00	13.00
	23	102	10.99	1.69	5.00	13.00
	24					

^a All age-based forms showed significant age group increases in scores based on age in months, $p < 0.001$.

estimated for the South American sample using trend analysis and linear interpolation. Using the ogives for groups defined by half months, it was possible to estimate the score distribution for the missing half month and adjust the distribution appropriately. There was no need for adjustment when comparing the other five benchmark ages to the US sample since the age criteria were similar between the two groups.

2.6. Data analysis

The BINS item scores were summed to obtain the total score for each infant and infants were classified into risk categories (Low, Moderate, and High) based on the US norms. Next, differences at the item and test level by selected characteristics and similarity of the South American distributions to the US norms were evaluated.

Descriptive statistics for the total BINS score were calculated for the overall sample as well as by sex and language. Group differences were evaluated using *t*-tests. At the item level, differences in performance by language and sex were tested using logistic regression.

Differences in the distributions of scores in the South American sample and the US norming sample were conducted at six benchmark ages (3, 6, 9, 12, 18, and 24 months) corresponding to the six ages where data were collected for the US standardization.¹ Differences were tested using the Kolmogorov–Smirnov test. In addition, chi-square tests were used to compare the distributions of the risk categories (low, moderate, and high) between the two samples.

For all analyses, $\alpha = .01$ was chosen to balance the need to control Type I error given multiple tests and the desire to identify items or total scores that might potentially show group differences.

3. Results

3.1. South American BINS scores

Descriptive statistics by age group and specific age in months for the BINS scores in the South American sample are presented in Table 2. As expected for a sample of infants without major health conditions, the BINS scores were high. As can be seen in this table, with the exception of a slight decrease in the mean score at 22 months, within each age group the mean score increased steadily with the infants' age. All age-based forms showed significant ($p < 0.001$) increases in scores

¹ Other ages were not directly observed in the US sample and, at those ages, the BINS distributions were interpolated by the BINS test developers based on the data from benchmark ages.

Table 3
Distributions of South American BINS scores by sex and language.

Age group (# of BINS items)	Total sample Mean (SD) ^a [25th,75th percentile]	Females Mean (SD) ^a [25th,75th percentile]	Males Mean (SD) ^a [25th,75th percentile]	Spanish Mean (SD) ^a [25th,75th percentile]	Portuguese Mean (SD) ^a [25th,75th percentile]
Age 3–4 months (11 items)	9.56 (1.12) [9,10] N = 389	9.59 (1.07) [9,10] N = 186	9.53 (1.16) [9,10] N = 203	9.62 (1.11) [9,10] N = 299	9.36 (1.11) [9,10] N = 90
Age 5–6 months (13 items)	10.86 (1.54) [10,12] N = 382	10.80 (1.39 ^b) [10,12] N = 191	10.92 (1.69 ^b) [10,12] N = 191	10.86 (1.58) [10,12] N = 303	10.86 (1.40) [10,12] N = 79
Age 7–10 months (13 items)	10.94 (1.65) [10,12] N = 487	11.01 (1.57) [10,12] N = 237	10.87 (1.72) [10,12] N = 250	10.89 (1.66) [10,12] N = 370	11.07 (1.60) [10,12] N = 117
Age 11–15 months (11 items)	9.62 (1.31) [9,11] N = 450	9.65 (1.38) [9,11] N = 230	9.60 (1.24) [9,10.5] N = 220	9.64 (1.32) [9,11] N = 341	9.56 (1.28) [9,11] N = 109
Age 16–20 months (11 items)	8.59 (1.64) [8,10] N = 377	8.91 ^b (1.58) [8,10] N = 195	8.25 ^b (1.65) [7,9] N = 182	8.65 (1.67) [8,10] N = 280	8.42 (1.57) [7,10] N = 97
Age 21–24 months ^c (13 items)	10.51 (2.03) [9,12] N = 386	10.76 ^b (2.00) [10,12] N = 190	10.27 ^b (2.03) [9,12] N = 195	10.57 (2.04) [9,12] N = 300	10.33 (1.98) [9,12] N = 86

^a Standard deviations of the mean BINS scores are in parentheses. The 25th and 75th percentiles are in brackets []. *N* indicates the number of children in each subgroup.

^b For 5–6 months, male scores were significantly more variable than female scores, $\alpha = 0.05$. For 16–20 months and 21–24 months, females had significantly higher scores (performed better) than males, $\alpha = 0.05$.

^c One child in the 21–24 months group was missing data on sex.

based on age in months. For example, in the 7–10 month age group, the mean BINS score was 9.99 at 7 months, 10.44 at 8 months, 11.42 at 9 months, and 11.66 at 10 months.

Table 3 reports BINS scores by sex and language. Differences were compared for each age group between male and female infants and by language. At most age levels, female infants performed higher than male, however the difference was only significant for the older age groups (16–20 months and 21–24 months). In addition, scores of male infants were more variable than those of females, however, the variability difference was only significant at 5–6 months. When comparing the BINS scores by language (Portuguese versus Spanish), none of the comparisons yielded significant difference.

3.1.1. Item analysis

Sex and language differences were also examined at the item level using logistic regression. For each item, performance differences based on language, sex, and the interaction of sex and language were tested. Because of the large number of tests being conducted, an alpha of .01 was chosen to limit the number of incorrectly identified comparisons to 1% as described above. The BINS scores were significantly different only for 2 items. Language had a significant main effect on item 1 (an item that measures conjugate gaze while looking at a small object) of the 5–6 month BINS form. A closer look at the data showed that in Spanish speaking countries, 94.2% of the babies demonstrated this skill versus 82.5% in Brazil (Portuguese-speaking). Both sex and the interaction between sex and language had significant effects on item 4 (an item that measures object permanence) of the 7–10 month BINS form. A lower percentage of female infants from Brazil demonstrated this skill than males from Brazil or infants of either sex from the Spanish-speaking countries. Given the large number of tests conducted for this analysis (more than 200), the observed result of 3 significant tests is in line with the chosen alpha (.01). Detailed results for the item-level logistic regression are available from the authors upon request.

3.2. Comparison of BINS South American scores with US norms

At six benchmark ages, the six observed ages in the US standardization sample (3, 6, 9, 12, 18, 24 months), the distribution of the total BINS scores in the South American sample² was compared to the distribution in the US standardization sample (Table 4). None of the comparisons were significant at the chosen $\alpha = .01$, however, significant differences would have been found at three and six months had we used the more liberal $\alpha = .05$. Examining the differences showed that at three months, infants in the South American sample outscored infants in the US sample, whereas at six months the relationship was reversed. Appendix A provides the distribution of the total BINS scores in the South American sample compared to the distribution in the US standardization sample for the non-benchmark months.

We also compared the BINS risk category distributions, based on the BINS guidelines, between the US and South American samples for the six benchmark ages. Table 5 reports the percent of infants in each risk category along with the p -value for the chi-square test for risk category differences between the two populations. Only the results for the youngest age (3 months) were significant at $\alpha = .01$. Cell chi-square statistics indicate that there were fewer high-risk infants in the South American sample than would be expected had the two distributions been identical. Had we used the more liberal $\alpha = .05$, we would have also found significance at ages 6 and 12 months. There was no single explanation for the differences at these ages – at 6 months there were more infants in the moderate risk category in South America whereas 12 months was similar to 3 months in that there were fewer high-risk infants in the South American sample.

4. Discussion

This study evaluated the utility of the BINS in South America by recruiting and screening a large sample of healthy infants in several South American countries. The study results indicate that the instrument appears to be both feasible and appropriate for neurodevelopmental screening in South America. Specific findings include differences in BINS scores by sex for certain age groups and overall lower rates of high-risk infants and higher BINS scores in the South American sample compared to the US standardization sample. Each of these findings is discussed below including their implication for the instrument's utility for South America.

Within the South American sample, we evaluated differences in BINS scores based on factors such as sex and language. As mentioned above, most differences by sex were statistically insignificant. The trend in three of the age groups, particularly the older two groups, was for females to score higher than males. Consistency in findings across sexes along with the lack of language differences in BINS scores (Spanish versus Portuguese) suggests no biases that are related to these factors with using the BINS in South America.

Comparing the South American results with the US norms, we found significantly lower rates of at risk infants in the South American sample compared to the US sample. This may be due to the use of more stringent inclusion criteria for healthy children in the South American sample. The South American children were identified by physicians and

² To account for an incorrect age definition at 3 months, an adjusted BINS score distribution was estimated for the South American sample using trend analysis and linear interpolation.

Table 4
Cumulative percentage at or above the BINS total score in the US and South American samples.

Age (in months)	Total score	US (cumulative %) ^a	South America (cumulative %)	D_{\max} ($p > KS$)
3 ^b	11	12	6	.203 ($p < .014$)
	10	33	41	
	9	58	79	
	8	80	92	
	7	89	98	
	6	95	99	
	5	98	100	
	4	100	100	
	3 and below	100	100	
6	13	32	16	.164 ($p < .048$)
	12	54	44	
	11	72	73	
	10	85	90	
	9	93	99	
	8	96	100	
	7	98	100	
	6	99	100	
	5 and below	100	100	
9	13	19	20	.067 (n.s.)
	12	58	51	
	11	79	79	
	10	89	93	
	9	96	99	
	8	98	99	
	7	98	100	
	6	99	100	
	5 and below	100	100	
12	11	14	15	.171 ($p < .081$)
	10	39	49	
	9	62	79	
	8	74	91	
	7	90	98	
	6	94	99	
	5	100	100	
	4	100	100	
	3 and below	100	100	
18	11	7	11	.080 (n.s.)
	10	36	28	
	9	48	51	
	8	70	72	
	7	82	89	
	6	91	98	
	5	93	100	
	4	95	100	
	3	97	100	
	2 and below	100	100	
24	13	23	21	.153 ($p < .188$)
	12	42	40	
	11	59	70	
	10	69	84	
	9	79	92	
	8	90	95	
	7	92	98	
	6	99	99	
	5 and below	100	100	

^a Data on US infants are from the BINS manual.

^b Data on South American infants were primarily from babies 3 to 3.5 month old infants. Linear interpolation was used to better approximate the distribution that would have been obtained if a sample of 2.5–3.5 month old infants had been obtained.

included only infants who were attending routine well-child visits. The US sample was a broader community based sample. Some of the exclusion criteria in the South American sample were not explicitly used in the US standardization sample. The US criteria for the nonclinical sample included normal length of gestation (38–42 weeks) and no prenatal, perinatal, or neonatal medical complications. Minor perinatal incidents were allowable such as jaundice that was treatable by phototherapy and required no extended hospital stay (Aylward, 1995). Therefore, it is likely that the South American

Table 5
Percent of infants by risk category in the US and South American samples.

Age (in months)	Risk category	US (%)	South American (%)	χ^2
3 ^a	Low	58	78	$p < 0.001$
	Moderate	31	20	
	High	11	2	
6	Low	54	44	$p < 0.039$
	Moderate	31	46	
	High	15	10	
9	Low	58	51	$p < 0.187$
	Moderate	31	42	
	High	11	7	
12	Low	39	49	$p < 0.015$
	Moderate	51	49	
	High	10	2	
18	Low	48	51	$p < 0.121$
	Moderate	43	47	
	High	9	2	
24	Low	59	70	$p < 0.203$
	Moderate	31	25	
	High	10	5	

^a South American distribution involved interpolating data for infants aged 2.5–3.0 month olds and combining that distribution with the observed distribution for 3.0–3.5 month olds.

sample included healthier children than the US standardization sample, which feasibly may have resulted in the better scores.

Based on this large, multi-national sample, it appears that the BINS can be applied to a population of healthy infants for neurodevelopmental screening, regardless of whether it is translated into Spanish or Portuguese. The BINS items evaluate fairly basic, generic developmental tasks that seem to be insensitive to cultural and socioeconomic differences. This wide range of application is probably influenced by the emphasis on neurodevelopment, e.g. tone and motor. The content is perhaps less affected by language and cultural factors than would other screening instruments that have more complex tasks or reliance on verbal factors. However, noteworthy is the fact that language items are also included in the BINS, particularly from 7 to 10 months onward, and involve making sounds, recognizing words, following commands, naming pictures and objects, pointing to pictures and a doll's body parts, word combinations, and gesturing. The study physicians found the BINS items to be culturally appropriate for South America. Further studies with diagnostic assessments of these children are needed to directly evaluate the sensitivity and specificity of the BINS in this population.

This is a first step in evaluating the efficacy of using the BINS in South America. There are limitations to this study that need to be noted. While the study sample has significant geographic and socioeconomic diversity, which considerably enhanced its representativeness, it nonetheless was purposefully selected. Future studies that include random and population-based samples of infants, although difficult to obtain, would be helpful to better evaluate the generalizability of the results to the population of infants in the study countries. The current study included only a healthy sample; application to a clinical sample is needed to further assess the psychometric properties of the BINS in South America.

5. Conclusions

In conclusion, the BINS is a useful instrument for neurodevelopmental screening in South America, with no apparent systematic biases or feasibility limitations. Further studies are needed to confirm the BINS utility in South America, including evaluating its properties among children with health problems and its concurrent validity with other diagnostic instruments.

Acknowledgements

The authors thank the pediatricians, study coordinators, staff, and families in South America who made this research possible. The outcome data in this work are derived from the Bayley Infant Neurodevelopment Screener. Copyright © 2004. Harcourt Assessment Inc. Used with Permission. All rights reserved.

Appendix A. Cumulative percentage at or above the BINS total score in the US and South American samples for non Benchmark months.^a

Age (in Months)	Total Score	US (Cumulative %) ^b	South American (Cumulative %)
4	11	26	28
	10	50	63
	9	65	86
	8	84	97
	7	90	99
	6	98	100
	5	99	100
	4	100	100
	3 & below	100	100
5 ^c	13	17	5
	12	48	27
	11	60	51
	10	73	71
	9	78	88
	8	86	93
	7	90	96
	6	94	98
	5	96	99
	4	98	100
	3	98	100
	2	99	100
	1 & below	100	100
7 ^c	13	17	6
	12	45	22
	11	65	36
	10	79	62
	9	86	81
	8	94	88
	7	97	94
	6	98	98
	5	100	99
	4 & below	100	100
8	13	27	9

Appendix A (Continued)

	12	51	27
	11	71	57
	10	84	72
	9	89	87
	8	95	97
	7	97	98
	6	99	98
	5	100	98
	4 & below	100	100
10	13	33	32
	12	65	59
	11	79	86
	10	90	91
	9	94	97
	8	99	100
	7	100	100
	6	100	100
	5 & below	100	100
11^c	11	18	4
	10	43	32
	9	54	57
	8	66	82
	7	76	90
	6	89	100
	5	93	100
	4	98	100
	3 & below	100	100
13	11	25	26
	10	45	59
	9	60	80
	8	70	92
	7	82	96
	6	94	99
	5	96	100
	4	99	100
	3	100	100
	2 & below	100	100
14	11	27	40
	10	49	71

Appendix A (Continued)

	9	62	93
	8	72	96
	7	84	99
	6	94	99
	5	97	100
	4	100	100
	3	100	100
	2 & below	100	100
15	11	30	62
	10	50	86
	9	68	94
	8	77	98
	7	88	99
	6	95	100
	5	98	100
	4	100	100
	3	100	100
	2 & below	100	100
16 ^c	11	9	1
	10	21	6
	9	38	22
	8	51	50
	7	63	63
	6	73	81
	5	84	91
	4	88	97
	3	96	100
	2	98	100
	1	98	100
	0	100	100
17	11	10	3
	10	26	23
	9	42	56
	8	53	73
	7	67	90
	6	77	98
	5	88	100
	4	91	100
	3	96	100

Appendix A (Continued)

	2	98	100
	1	98	100
	0	100	100
19	11	15	15
	10	37	39
	9	52	67
	8	66	85
	7	77	90
	6	85	96
	5	94	98
	4	95	100
	3	96	100
	2	98	100
	1	99	100
	0	100	100
20	11	19	20
	10	41	49
	9	56	71
	8	67	89
	7	79	95
	6	87	96
	5	94	99
	4	96	99
	3	96	100
	2	98	100
	1 & below	100	100
21^c	13	15	4
	12	33	18
	11	42	52
	10	59	64
	9	66	77
	8	73	88
	7	83	88
	6	90	97
	5	93	98
	4	96	99
	3	98	100
2 & below	100	100	
22	13	17	9

Appendix A (Continued)

	12	38	25
	11	47	47
	10	61	60
	9	71	77
	8	78	86
	7	85	90
	6	91	96
	5	95	100
	4	97	100
	3	99	100
	2 & below	100	100
23	13	22	21
	12	39	40
	11	52	70
	10	64	84
	9	73	92
	8	81	95
	7	88	98
	6	95	99
	5	96	100
	4	98	100
	3	100	100
	2 & below	100	100

^aGray shading is low risk and pink shading is high risk.

^bUS distributions are based on interpolation from the data collected at 3, 6, 9, 12, 18, and 24 months.

^cData on South American infants were primarily from older babies in this age range. Linear interpolation was used to better approximate the distribution that would have been obtained if the sample had been balanced.

References

- Anderson, P. J., & Doyle, L. W. (2008). Cognitive and educational deficits in children born extremely preterm. *Seminars in Perinatology*, 32., 51–58.
- Aylward, G. P., & Verhulst, S. J. (2008). Comparison of caretaker report and hands-on neurodevelopmental screening in high-risk infants. *Developmental Neuropsychology*, 33., 124–136.
- Aylward, G. P., & Verhulst, S. J. (2000). Predictive utility of the Bayley Infant Neurodevelopmental Screener (BINS) risk status classifications: Clinical interpretation and application. *Developmental Medicine and Child Neurology*, 42., 25–31.
- Aylward, G. P. (1998). Alternative risk grouping method for the Bayley Infant Neurodevelopmental Screener (BINS) in prediction of later dysfunction. *Journal of Developmental and Behavioral Pediatrics*, 19., 386–387.
- Aylward, G. P. (2010). The Bayley Infant Neurodevelopmental Screener (BINS): Different test and different purpose. In L. G. Weiss, T. Oakland, & G. Aylward (Eds.), *Bayley-III clinical use and interpretation*. Burlington, MA: Academic Press/Elsevier.
- Aylward, G. P. (1995). *The Bayley Infant Neurodevelopmental Screener manual*. San Antonio, TX: The Psychological Corporation.
- Aylward, G. P. (1994). Update on early developmental neuropsychologic assessment: The Early Neuropsychologic Optimality Rating Scales (ENORS). In M. G. Tramontana, & S. R. Hooper (Series Eds.) & *Advances in child neuropsychology* : Vol. 2. (pp. 172–200). Springer-Verlag: Berlin.
- Bayley, N. (1993). *Bayley Scales of infant development* (2nd ed.). San Antonio, TX: The Psychological Corporation.
- Benish, J. K. (1998). Review of the Bayley Infant Neurodevelopmental Screener. In *The thirteenth mental measurements yearbook*. Buros: Test Reviews Online.
- Blackman, J. A. (1999). Developmental screening: Infants, toddlers, and preschoolers. In M. D. Levine, W. B. Carey, & A. C. Crocker (Eds.), *Developmental-behavioral pediatrics* (3rd ed., Vol. 2, pp. 689–695). Philadelphia, PA: Saunders.
- Boulet, S. L., Boyle, C. A., & Schieve, L. A. (2009). Health care use and health and functional impact of developmental disabilities among US children, 1997–2005. *Archives of Pediatrics and Adolescent Medicine*, 163., 19–26.
- Boulet, S. L., Schieve, L. A., & Boyle, C. A. (2009). Birth weight and health and developmental outcomes in US children, 1997–2005. *Maternal and Child Health Journal*, (November) [Epub ahead of print].
- Castilla, E. E., & Orioli, I. M. (2004). ECLAMC: The Latin-American collaborative study of congenital malformations. *Community Genetics*, 7., 76–94.
- Constantinou, J. C., Adamson-Macedo, E. N., Mirmiran, M., Ariagno, R. L., & Fleisher, B. E. (2005). Neurobehavioral assessment predicts differential outcome between VLBW and ELBW preterm infants. *Journal of Perinatology*, 25., 788–793.
- Council on Children with Disabilities. (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*, 118., 405–420.
- Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, 47., 87–122.
- Dobrez, D., Sasso, A. L., Holl, J., Shalowitz, M., Leon, S., & Budetti, P. (2001). Estimating the cost of developmental and behavioral screening of preschool children in general pediatric practice. *Pediatrics*, 108., 913–922.
- Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 Cohort. *Economica*, 70., 73–97.
- Govindaraj, R., Chellaraj, G., & Murray, C. J. (1997). Health expenditures in Latin America and the Caribbean. *Social Science and Medicine*, 44., 157–169.

- Gucuyener, K., Ergenekon, E., Soysal, A. S., Aktas, A., Derinoz, O., Koc, E., et al. (2006). Use of the Bayley Infant Neurodevelopmental Screener with premature infants. *Brain Development*: 2., 104–108.
- Individuals with Disabilities Education Act, 20 U.S.C § 1400, et. seq. (1997).
- Individuals with Disabilities Education Improvement Act of 2004. HR 1350, November 19, 2004.
- Jelliffe-Pawlowski, L. L., Shaw, G. M., Nelson, V., & Harris, J. A. (2003). Risk of mental retardation among children born with birth defects. *Archives of Pediatric and Adolescent Medicine*: 157., 545–550.
- Leonard, C. H., Piecuch, R. E., & Cooper, B. A. (2001). Use of the Bayley Infant Neurodevelopmental Screener with low birth weight infants. *Journal of Pediatric Psychology*: 26., 33–40.
- López-Camelo, J. S., Campaña, H., Santos, R., & Poletta, F. A. (2006). Effect of the interaction between high altitude and socioeconomic factors on birth weight in a large sample from South America. *American Journal of Physical Anthropology*: 129., (2), 305–310.
- Macias, M. M., Saylor, C. F., Greer, M. K., Charles, J. M., Bell, N., & Katikaneni, L. D. (1998). Infant screening: The usefulness of the Bayley Infant Neurodevelopmental Screener and the Clinical Adaptive Test/Clinical Linguistic Auditory Milestone Scale. *Journal of Developmental and Behavioral Pediatrics*: 19., 156–161.
- Paxson, C., & Schady, N. (2007). Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. *Journal of Human Resources*: 42., 49–84.
- Wehby, G. L., Castilla, E. E., Goco, N., Rittler, M., Cosentino, V., Javois, L., et al. (2006). Description of the methodology used in an ongoing pediatric care interventional study of children born with cleft lip and palate in South American [NCT00097149]. *BMC Pediatrics*: 6., 9.
- Wehby, G. L., Castilla, E. E., & Lopez-Camelo, J. (2010). The impact of altitude on infant health in South America. *Economics and Human Biology*: 8., 187–211.
- Wehby, G. L., Murray, J. C., Castilla, E. E., Lopez-Camelo, J. S., & Ohsfeldt, R. L. (2009). Quantile effects of prenatal care utilization on birth weight in Argentina. *Health and Economics*: 18., 1307–1321.