

---

## The reliability and validity of Bayley-III cognitive scale in China's male and female children

### ABSTRACT

**Background:** Since publication in 2006, the Bayley-III scale has been used widely in pediatric populations worldwide; however, there have been very few studies which examined the usefulness and the potential sex differences in a Chinese context.

**Aims:** To assess the reliability and validity of the Bayley-III cognitive scale, and detect possible sex differences in term children so as to provide evidence for clinical and research use in China.

**Study design:** Cross-sectional study.

**Participants and outcome measures:** Of the 1589 children from 3 healthcare institutions that were initially recruited, a total of 1444 children were included in the final analysis. We randomly selected 5-10% children from the total sample to evaluate the test-retest, inter-rater and criteria-related reliability in order to meet the psychometric criteria of Bayley-III scale. Inter-item consistency, test-retest and inter-rater reliability of the scale were estimated using Split-half method and Intra-class Correlation Coefficient (ICC). The content validity was evaluated by the Item-level Content Validity Index (I-CVI). The Mann-Kendall trend test was performed to assess trends of cognitive development, and post-hoc Least Significant Difference test was used to detect age-appropriateness of items.

**Results:** Six developmental pediatricians were trained to administer the Bayley-III cognitive scale. Inter-item consistency (n=1444) with Guttman split-half coefficient was above 0.8, while test-retest (n=144) and inter-rater reliability (n=74) had good to excellent ICCs of over 0.9. The criteria-related validity (n=74) of Bayley-III was acceptable, and associations with Gesell Developmental Schedules (GDS) were mainly above 0.8. The raw score of Bayley-III scale in total subjects (n=1444) showed an increased trend across all months of age ( $p < 0.05$ ), and only the score in age group of 35M16D to 36M15D declined in females ( $p < 0.05$ , n=722). Female children presented a higher score than male children in all subjects and in the 18–23 months age group ( $p < 0.05$ ).

---

**Conclusions:** Our findings provide important evidence ~~for that the Bayley-III cognitive scale is a valid measurement which could be used in Chinese population—the adjustment of age-inappropriate items in Bayley-III cognitive scale~~, as well as the consideration of sex differences when used in a Chinese context.

**Keywords**

Bayley-III; Cognitive scale; Usefulness; Sex difference; China

---

## 1. Introduction

The Bayley Scales of Infant and Toddler Development-3rd Edition (Bayley-III) is a well-known psychometric instrument, which provides useful information for the early identification of infants who might have developmental problems [1]. Since publication in 2006, the Bayley-III has been used widely in pediatric populations worldwide [2-11], and numerous researchers have confirmed the strong reliability and validity of the instrument [9-15]. The instrument enables the assessment of cognition, receptive communication, expressive Communication, Fine Motor Development, and Gross Motor Development of infants and toddlers from 1 to 42 months old, with five separated subscales accordingly. Compared with its previously published edition, one of the significant changes of Bayley-III is that the Cognitive Scale was created in an attempt to isolate cognitive skills from language skills [16, 17], which is important for the assessment on very young children, especially infants and toddlers with very limited languages skills. The Bayley scales have been used in various countries such as the Netherlands [18], Iran [19], Australia[20] and Asian countries [21, 22]. However, cross-culture difference has been reported [22] in different countries [13, 18, 23, 24]. In children with developmental disorders such as Autism Spectrum Disorder(ASD) [25, 26] and Attention and Attention Deficit Hyperactivity Disorder (ADHD)[27, 28], the cognitive level in their early childhood is usually one of the predictive indicators for their final diagnosis; and any difficulty in cognitive development can have a strongly negative impact on their daily life. However, careful evaluation is required when cross-culture differences are considered based on cognitive

---

development, and differences in cognitive developmental appear more difficult to be expected between different cultures [24]. Therefore, in the present study, we focused on the Cognition cognition subscale in Bayley-III to explore its suitability on Chinese children. Between different cultures, the implicit assumptions about values, knowledge and communication are inherent in the whole concept of developmental and psychological testing. Thus, cultural differences might affect test performance and scores, especially where item suitability and sequence has not been adapted to the local culture. In a longitudinal study, the cognitive level [24] at different developmental stages showed a wide gap between two cultural environments when measured by Bayley-III. In contrast to the United States, early childhood in China is characterized by vastly different childcare practices and different exposure to educational games, books, toys, and multimedia on television and computers. Therefore the item suitability and testing sequence of Bayley-III may be distinct when using this scale on children from a Chinese population.

The other focus of the current study is to determine whether there are potential sex differences in early infancy between different cultures. Multiple differences between males and females, both in normal physiology and pathophysiology of diseases have been recently reported [29-32]. Histomorphometric studies confirmed a sexual dimorphism in human cerebral cortex, as males show higher average neuronal density, but have smaller neuronal units than females [33-35], which may have a potential influence on children's neuropsychological function. Moreover, in traditional low-income environments such as rural China, families may emphasize the development of human skills that are more in favor of sons at the expense of daughters [36]. This could help explain the sex differences of psychomotor

---

development in children, which may also influence the suitability and testing sequence of Bayley-III items on different sexes when used in a Chinese context. Unfortunately, there have been very few studies which take into account these potential sex differences when Bayley-III is used in a different culture context.

In this study, we aimed to 1) explore the reliability and validity of Bayley-III cognitive scale in term children across all ages as provided by the instrument. The objectives of this study were thus to analyze the reliability and validity of the Bayley-III cognitive scale in term children across all months of age in Mainland China, 2) explore the sex differences between male and female children when assessing the construct and known-group validity so as to provide evidence for ~~adjusting age inappropriate items~~ developing specific norms in Chinese children for clinical and research use.

## **2. Methods**

### *2.1 Participants*

We conducted a cross-sectional study in mainland China from May to December of 2011. We used a stratified sampling technique, with area, sex, and months of age as stratification variables. A total of 1589 children aged between 16 days to 42 months were selected from 3 children's healthcare institutions in medium-sized cities distributed across 3 geographic regions: North China, Middle China and East China. The months of age were selected based on the categories proposed in the Bayley-III technical manual (total of 48 age bands). The inclusion criteria included: children who were born at term, born without significant medical complications, did not have a history of medical complications, and were not currently

---

diagnosed with or receiving treatment for mental, physical or behavioral difficulties. The exclusion criteria included: presence of confounding conditions or developmental risk factors such as hearing or vision impairment~~without normal hearing or vision~~, taking medications that could affect performance or were admitted to hospital at the time of testing, and other nutrition, sleeping or infection problems during the clinical visit.

Formatted: Font: (Asian) +Body Asian (SimSun), 11 pt, Complex Script Font: 11 pt

Of the 1589 eligible children who were recruited, a total of 1444 children were included in the study (Figure 1). All information was kept confidential and was only accessible to the researchers. The study was approved by the Ethic Committee of Soochow University Children's Hospital (201101), China. Written informed consent was obtained from the parents or legal guardians of the participants prior to the questionnaire survey. Oral parental consents were obtained prior to conducting the investigation and tests.

## 2.2 Procedure

The survey was conducted during the well-child visits in the participating children's healthcare institutions. Nurses who took part in the check in and physical examination (weight, height and head circumference) were responsible for handing out the questionnaires to the children's parents. Parents of participating children were asked to fill out the questionnaires according to the attached illustration. Six developmental pediatricians were trained to administer the Bayley-III cognitive scale. The testers became familiar with the examination according to the test manual by carrying out a series of practice assessments on several children who did not take part in the study. Any problems associated with test administration during the training period were clarified by the administrator of this study prior

Formatted: Font: (Asian) +Body Asian (SimSun), 11 pt, Complex Script Font: 11 pt

---

to the test. The test environment was quiet and non-interfering, and all infants were required to be stable, sober and satiated. The testers encouraged the infants and toddlers to display their highest level of ability during the test. A trained pediatrician took the responsibility for conducting the entire test for each child in order to maximize both interpretation validity and assessment reliability.

Additionally, we randomly selected 5-10% children from the total sample in to evaluate the test-retest, inter-rater and criteria-related reliability in order to meet the psychometric criteria of Bayley-III. One hundred and forty four children across all months of age were randomly selected from the 1444 subjects who took part in the Bayley-III cognitive scale by the same tester (not the observed one) twice a week apart, in order to minimize drawing conclusions due to age-related changes in performance. Seventy-four children were randomly selected from all recruitment institutions and were tested by one tester while the other tester observed. The testers reversed their roles when half of the children completed the Bayley-III cognitive scale. Both testers scored each child's cognitive performance simultaneously, but independently. A score of homogeneity or consensus is given by judges [using a correlation analysis](#), and if various raters do not agree, then either the scale is defective or the raters needed to be re-trained. For assessment of the criteria-related validity, 184 children attended the GDS test a week after the second test of Bayley-III cognitive scale.

### *2.3 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. We first obtained formal permission to translate and validate the Bayley-III scale from the publishers (Pearson). We then started developing a Chinese version of Bayley-III, following the recommendations of Hambleton and Patsula (1999) and Herdman, Fox-Rushby and Badia (1998) for the translation and adaptation of a test, taking into consideration conceptual, item, semantic, operational, measurement and functional equivalences. Each step of this process was presented in the results section. The Chinese version of the Bayley-III was translated by a native Chinese speaker and independent professional who adapted the items into context and culture, and subsequently retranslated into English by two native English speakers who were blinded to the original version. The test manual and materials ~~will~~utilized the same trademark, logo, and design as used on the English version of the test. In this study, we explored the reliability and validity of the Bayley-III cognitive scale. Further study on other sub-scales of Bayley-III will be conducted in the near future.

The Gesell Developmental Schedules (GDS) was designed to provide a neurologic and intellectual assessment in infants and toddlers [37, 38]. The items are grouped into five main categories of functioning: the gross and fine motor skills, language development, adaptive behavior, and personal social behaviors. The results were expressed as developmental scores for the five categories. The Chinese version of the GDS was translated and revised by the Beijing Mental Development Cooperative Group (1985). It has been used widely in China to assess infant intellectual development for abnormality [39-41].

---

#### 2.4 Statistical analysis

The inter-item consistency was performed using Guttman split-half method. Coefficients above 0.75 referred to good reliability, and those above 0.5 were acceptable, while those below 0.5 suggested poor reliability [42]. The Intra-class Correlation Coefficient (ICC) was used to examine both the inter-rater and test-retest reliability based on the raw score and composite score of the scale. ICCs above 0.6 referred to good reliability, those between 0.2 and 0.5 meant moderate reliability, and those below 0.2 suggested poor reliability[43].

We used GDS to assess the criteria-related validity using the Pearson correlation analysis because there were only a small number of measurements with Chinese norm available on the age range we focused on. Correlation coefficients above 0.70 were considered strong, below 0.30 indicated weak, and those between 0.30 and 0.70 were moderate [44]. The known-group validity of the Bayley-III cognitive scale was assessed by comparing the composite scores between males and females at each age band using the two-dependent sample t-test.

To evaluate the construct validity of the Bayley-III cognitive scale, the Mann-Kendall trend test was performed to assess trends of the scale's raw score across all months of age. The Mann-Kendall trend test was a nonparametric test based on the Kendall rank correlation between the value of interest and time, which was used by Davenport in their study of the historical trend of child mortality [45]. Additionally, because the study design was cross-sectional across the months of age in infants and toddlers (not a repeated measurement data), we used a post-hoc Least Significant Difference (LSD) test to compare the means of raw scores between two adjacent age groups based on the categories proposed in the Bayley-III

---

technical manual, so as to provide the evidence for detecting any items which may not be suitable based on the months of age in Chinese children.

### **3. Results**

#### *3.1 The socio-demographic characteristics of the subjects*

Of the 1589 participants who were recruited in our initial study, 1444 children were included in the final study. Of the 1444 children, 722 were male (50.0%) and 722 were female (50.0%). These subjects were distributed almost equally across 48 age bands according to the Bayley-III technical manual. The mean age of the subjects was 16.14 months of age, with a standard deviation of 12.05. The socio-demographic characteristics of the subjects were shown in the Table 1. Differences between male and female children included head circumference, kaup index (a measurement of obesity based on height and weight that applies to children below 8 years old), and mother's vocation ( $p < 0.05$ ). However, there were no other significant differences related to socio-demographic characteristics between male and female children ( $p > 0.05$ ; Table 1).

#### *3.2 Reliability*

Inter-item consistency: The Guttman split-half coefficient of Bayley-III cognitive scale was 0.803, suggesting that the Inter-item consistency of the questionnaire was good.

Test-retest and inter-rater reliability: The study showed that ICCs for test-retest reliability (Table 2) raw score and composite score (the interval time of two tests were 1 weeks) were excellent, with all items above 0.9. ICCs for most of items, subscales and total score were above 0.9 (Table 2). The result also showed that the test-retest reliability of Bayley-III cognitive scale

---

was excellent, with an ICC above 0.9. Furthermore, inter-rater reliability of raw score and composite scores were also good, with an ICCs above 0.9 for both (Table 2).

### 3.3 Validity

Criteria-related Validity: There were strong associations between the raw scores of Bayley-III cognitive scale and gross motor skills, fine motor skills, adaptive behavior and language of GDS with correlation coefficients above 0.8 for all categories (each  $p < 0.05$ ; table 3), which illustrated the acceptable validity of the scale. The Bayley-III cognitive scale was only weakly associated with social skills of GDS with a correlation coefficient of less than 0.3 ( $p < 0.05$ ; table 3).

Construct validity: The raw score of Bayley-III cognitive scale increased gradually across the months of age based on the categories proposed in the Bayley-III technical manual (Mann-Kendall trend test:  $p < 0.05$ ; Figure 2). A similar trend was also found in male and females (each  $p < 0.05$ ; Figure 2). The mean and standard deviation of each month age are shown in Figure 2-4. However, we observed a decreased mean score in several age groups compared with their adjacent groups (previous) in Figure 2-4. However, in the sample of all subjects, the reduction in mean scores in age groups of 30M16D to 31M15D( $n=30$ ), and 35M16D to 36M15D( $n=30$ ) when compared with their adjacent previous groups were not statistically significant (the values of  $p$  were 0.23 and 0.74 respectively) (Figure 2). In males, we found that the reduction in mean scores in age groups of 30M16D to 31M15D( $n=15$ ), 32M16D to 33M15D( $n=15$ ), and 36M16D to 39M15D( $n=15$ ) ( $p$  values of 0.24, 0.55 and 0.94, respectively; Figure 2), which were not statistically significant. Interestingly, in females, we found that the

---

mean scores in age groups of 35M16D to 36M15D(n=15) were decreased significantly compared with its adjacent previous group ( $p=0.02$ ). However, in other age groups of 23M16D to 24M15D (n=15), 29M16D to 30M 15D(n=15), ~~and 30M16D to 31M15D(n=15),~~ the reduction in mean scores were not statistically significant ( $p$  values of 0.20, 0.53 and 0.58, respectively; Figure2).

Known-group Validity: Female children presented generally higher composite scores than male children based on the total score from 0 to 42 months of age ( $p < 0.05$ ; Table 4). However, when we divided the subjects into seven age bands, there were differences in both raw and composite scores between male and female children at 18–23 months of age ( $p < 0.05$ ; Table 4).

#### 4. Discussion

Our study sheds light into the suitability of applying Bayley-III cognitive scale in a Chinese population. To our knowledge, this is the first study which explored the reliability and validity of Bayley-III scale across children of all months of ages in Mainland China, with a specialized focus on sex difference. Our results showed a good to excellent reliability and validity of Bayley-III cognitive scale, and found that girls generally performed better than boys on the Bayley-III cognitive scale according to its total score across all months of age. Additionally, we observed that the test score decreased with ages in some group of Chinese females, which suggested a slightly different developmental trajectories concerning cognitive development in females. ~~been inappropriate with regards to the testing sequence based on American norms, which suggested a need to adjust the testing sequence of items when using Bayley III in Chinese~~

---

**children.** Further, sex differences in the scale may also need to be considered in future studies.

The inter-item consistency assesses whether parts of an assessment are in fact measuring something similar to what the whole assessment claims to measure [46]. According to our results, the Chinese version of Bayley-III cognitive scale seems to be reliable in a Chinese population when considering its internal consistency compared with the original and other versions in different languages. The strong test-retest reliability (ICC) of the scale indicated that the results were stable over time (did not vary at one week intervals), which was consistent with those in the English version of Bayley-III scale, as well as those used in other countries [6] and listed in the technical manual [47]. Our study also indicated a good inter-rater reliability. These results may be due to the fact that testers had received prior training on the application of the Bayley-III cognitive scale, which can enable the reliable assessment of term children in China. Our results are consistent with previous reports for the instrument in Taiwanese children at 4 to 24 months of age [14], English children between 29 to 41 months of age [48], and Nepalese children across all months of age (1-42month) [9].

In our current study, the selected aspects of validity for Bayley-III cognitive scale involved the criteria-related, construct, and known-group validity. Evidence of criterion-related validity of Bayley-III cognitive scale was provided by the comparison with the GDS used in this study. Both scales cover a wide variety of skills organized in a developmental sequence, in order to identify infants and toddlers with developmental delay. In our study, the Bayley-III cognitive scale (raw score) was highly correlated with the gross motor skills, fine motor skills, adaptive, and language of **the** GDS (>0.8), and was weakly associated with social skills. This result is

---

inconsistent with results from an American population, and as reported in the technical manual of Bayley-III, the scale was not strongly correlated to any of the comparison tests. It had been reported that the cognitive level during infancy and toddlerhood was closely related with their performance of motor and language [49-51]. However, the strong correlations of cognitive scale with the GDS indicate that this cognitive scale of Bayley-III may have actually measured something else (the gross motor skills, fine motor skills, adaptive, and language) ~~other than those other items~~. Therefore, the conflicting results in our study in relation to the technical manual may need to be confirmed in further studies [47].

The construct validity is also important for deciding the suitability of Bayley-III cognitive scale. We observed an increasing trend of children's cognitive performance across all months of age based on the categories proposed in the technical manual of Bayley-III. When we further compared the mean of raw scores between two adjacent age groups, the results showed the raw score of Bayley-III cognitive scale increased gradually in infants. However, we found that the raw score had actually decreased over some adjacent older age bands ~~between a few comparisons~~. This may be explained by the differences in child-rearing practices under different cultures, which can hasten or delay the onset of specific skills that are measured using the Bayley-III scale [52]. Furthermore, experiences with different types of environments or toys can affect how children perform on specific items [24]. Interestingly, the statistically significant findings were only found in females, where the score declined in age group of 35M16D to 36M15D compared with its adjacent previous group. This suggests that sex differences should be considered when we evaluate the Bayley-III scale's suitability in China.

Formatted: Font: (Asian) +Body Asian (SimSun), 11 pt, Complex Script Font: 11 pt

---

In order to further differentiate the sex differences in Chinese children, we assessed the known-group validity of the Bayley-III cognitive scale in male and female children by comparing the raw and composite scores. Female children presented generally higher composite scores in all subjects aged from 0 to 42 months of age. Previous research has found that girls consistently outperformed boys in specific and general language performance in typical developing children [32], which significantly influenced the cognitive ability in toddlers. Such psychometric sex differences may reflect the structural cortical differences between males and females, where females were associated with the greater number of neuronal connections, which may result in the superior memory and language ability in females [35, 53].

—

Memory is one of the primary core cognitive abilities [54-57], and language performance also has a positive influence on cognitive ability in early childhood [24]. Additionally, we observed the sex differences in children aged 18–23 months when we divided the subjects into seven age bands. The age band of 18-23 months is one of the important stages for children in developing cognitive and language milestones [58, 59]. The sex difference may likely have emerged from this age band. Further research is needed to examine the stability of sex difference across different age bands. Moreover, our results disagreed with the assumption that Chinese families may emphasize the development of human skills that are more in favor of sons at the expense of daughters. A possible explanation may be that our sample of infants and toddlers are brought up in a more contemporary society, where the preference for sons over daughters has less of an

---

influence compared to older generations.

Additionally, the composite scores for the Bayley-III cognitive score in both male and female children were higher than the norms of mean (SD) score of 100 (15), and we did not find children with developmental delay (composite score less than 70) in our subjects, which is similar to previous findings that showed developmental outcomes were often overestimated and consequently this underestimated the disabilities in full term children [14]. Additionally, in our study, the majority of subjects were typically development children, which is different from the standardization sample listed in the technical manual. Approximately 10% of the standardization sample from American population included children selected from the special group studies with clinical disorders (e.g. Down Syndrome, cerebral palsy, pervasive developmental disorder, premature birth and at risk for developmental delay) [47]. Consequently, the inclusion of children with disabilities in the normative sample might lower the mean score, which may limit the tests' ability to diagnose children with mild disabilities.

Our study showed a good to excellent reliability and validity of Bayley-III cognitive scale.

Therefore, it is possible to use it for children aged 1-42 months in China.

Formatted: Font: Not Bold

Formatted: Font: Not Bold

However, a nation-wide survey with all subscales is still needed because of the socio-economic varieties in different areas of China. Other important findings were that girls generally

Formatted: Font: Not Bold

performed better than boys, and girls showed a slightly different developmental trajectory from

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light

boys on Bayley-III cognitive scale. Developing a sex-specific norm or a adjustment for

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light

---

~~cut-off score of the scale in different sex may deserve to be further considered in our further study in future studies. Unfortunately, we did not provide the information of which specific test items were inappropriate with regards to the testing sequence based on the American norms. In our study, however, it is the fact that At present, China still lack of valid the neurobehavioral measurements for children which could be compared with those in other nations. Therefore, it is important to preserve the international comparability of Bayley-III when it is used in Chinese population. Therefore, we should ~~change the sequence of items delete the age-inappropriate items or change the sequence of items with great caution, and may consider these changes based on a study with lager and national samples. This could potentially lead to a child with disabilities not being detected and not being referred to an appropriate intervention program in the clinical settings, if the testing sequence remained the same when applying the Bayley III scale on Chinese children. Moreover, the sex difference should be considered when we adjust the testing sequences and establish the Chinese norms in future studies. The identification of the sex-related psychomotor development could be of particular interest when further evaluation criterion and interventions are set up.~~ Additionally, s~~Similar to previous literature, our results also showed an overestimation of the developmental outcome and underestimation of the cognitive disability, however, these results might be due to differences in the samples between our study and the technical manual. A varying degree of upward adjustment of cut-off score of the scale is also recommended for the accurate identification of developmental delay.

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light, 11 pt, Font color: Auto, Complex Script Font: 11 pt

Formatted: Font: (Asian) Kepler Std Light

#### **Conflict of interest statement**

None declared.

---

## Acknowledgements

The authors wish to thank all researchers and medical practitioners who participated in this study for the distribution and collection of the questionnaires, as well as administrating the cognitive scale testing for children. We also thank Bing Wang for acquiring the Research Translation License Agreement of the Bayley-3 cognitive scale. This study was supported by National Natural Science Foundation of China (81673179), the Science and Technology Commission of Shanghai Municipality (18140903100), the Shengkang Hospital Development Center under Grant(SHDC12016239), the Fourth Round of Shanghai Three-year Action Plan on Public Health Discipline and Talent Program: Women and Children's Health under Grant (15GWZK0401), Shanghai municipal medical and health discipline construction projects under Grant (2017ZZ02015).

## References

- [1] Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatric research*. 2014;75:670-4.
- [2] Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW, Victorian Infant Collaborative G. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med*. 2010;164:352-6.
- [3] Msall ME. The Bayley-III scale underestimates developmental delay in extremely premature and extremely low birth weight infants. *J Pediatr*. 2010;157:863-4.
- [4] 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;128:e794-800.

- [5] Vohr B, McGowan E, McKinley L, Tucker R, Keszler L, Alksninis B. Differential Effects of the Single-Family Room Neonatal Intensive Care Unit on 18- to 24-Month Bayley Scores of Preterm Infants. *J Pediatr*. 2017.
- [6] Azari N, Soleimani F, Vameghi R, Sajedi F, Shahshahani S, Karimi H, et al. A Psychometric Study of the Bayley Scales of Infant and Toddler Development in Persian Language Children. *Iranian Journal of Child Neurology*. 2017;11:50-6.
- [7] Anderson PJ, Burnett A. Assessing developmental delay in early childhood - concerns with the Bayley-III scales. *The Clinical Neuropsychologist*. 2017;31:371-81.
- [8] Torras-Mana M, Gomez-Morales A, Gonzalez-Gimeno I, Fornieles-Deu A, Brun-Gasca C. Assessment of cognition and language in the early diagnosis of autism spectrum disorder: usefulness of the Bayley Scales of infant and toddler development, third edition. *J Intell Disab Res*. 2016;60:502-11.
- [9] Manandhar SR, Dulal S, Manandhar DS, Saville N, Prost A. Acceptability and Reliability of the Bayley Scales of Infant Development III Cognitive and Motor Scales among Children in Makwanpur. *Journal of Nepal Health Research Council*. 2016;14:47-50.
- [10] Veldhuizen S, Clinton J, Rodriguez C, Wade TJ, Cairney J. Concurrent validity of the Ages And Stages Questionnaires and Bayley Developmental Scales in a general population sample. *Acad Pediatr*. 2015;15:231-7.
- [11] Hanlon C, Medhin G, Worku B, Tomlinson M, Alem A, Dewey M, et al. Adapting the Bayley Scales of infant and toddler development in Ethiopia: evaluation of reliability and validity. *Child: care, health and development*. 2016;42:699-708.
- [12] Spencer-Smith MM, Spittle AJ, Lee KJ, Doyle LW, Anderson PJ. Bayley-III Cognitive and Language Scales in Preterm Children. *Pediatrics*. 2015;135:e1258-65.
- [13] Cromwell EA, Dube Q, Cole SR, Chirambo C, Dow AE, Heyderman RS, et al. Validity of US norms for the Bayley Scales of Infant Development-III in Malawian children. *European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society*. 2014;18:223-30.
- [14] Yu YT, Hsieh WS, Hsu CH, Chen LC, Lee WT, Chiu NC, et al. A psychometric study of the Bayley Scales of Infant and Toddler Development - 3rd Edition for term and preterm Taiwanese infants. *Res Dev Disabil*. 2013;34:3875-83.
- [15] Luttikhuisen dos Santos ES, de Kieviet JF, Konigs M, van Elburg RM, Oosterlaan J. Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early human development*. 2013;89:487-96.
- [16] Lowe JR, Erickson SJ, Schrader R, Duncan AF. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing? *Acta paediatrica*. 2012;101:e55-8.
- [17] Bos AF. Bayley-II or Bayley-III: what do the scores tell us? *Developmental medicine and child neurology*. 2013;55:978-9.
- [18] Steenis LJ, Verhoeven M, Hessen DJ, van Baar AL. Performance of Dutch children on the Bayley III: a comparison study of US and Dutch norms. *PloS one*. 2015;10:e0132871.
- [19] Azari N, Soleimani F, Vameghi R, Sajedi F, Shahshahani S, Karimi H, et al. A Psychometric Study of the Bayley Scales of Infant and Toddler Development in Persian Language Children. *Iranian journal of child neurology*. 2017;11:50-6.
- [20] Msall ME. Measuring outcomes after extreme prematurity with the Bayley-III Scales of infant and

---

toddler development: a cautionary tale from Australia. *Archives of pediatrics & adolescent medicine*. 2010;164:391-3.

[21] Godamunne P, Liyanage C, Wimaladharmasooriya N, Pathmeswaran A, Wickremasinghe AR, Patterson C, et al. Comparison of performance of Sri Lankan and US children on cognitive and motor scales of the Bayley scales of infant development. *BMC research notes*. 2014;7:300.

[22] Yu YT, Hsieh WS, Hsu CH, Chen LC, Lee WT, Chiu NC, et al. A psychometric study of the Bayley Scales of Infant and Toddler Development - 3rd Edition for term and preterm Taiwanese infants. *Research in developmental disabilities*. 2013;34:3875-83.

[23] Krogh MT, Vaever MS. Bayley-III: Cultural differences and language scale validity in a Danish sample. *Scandinavian journal of psychology*. 2016;57:501-8.

[24] Lohaus A, Keller H, Lamm B, Teubert M, Fassbender I, Freitag C, et al. Infant development in two cultural contexts: Cameroonian Nso farmer and German middle-class infants. *J Reprod Infant Psych*. 2011;29:148-61.

[25] Dawson G. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Dev Psychopathol*. 2008;20:775-803.

[26] Rogers SJ, Estes A, Lord C, Vismara L, Winter J, Fitzpatrick A, et al. Effects of a brief Early Start Denver model (ESDM)-based parent intervention on toddlers at risk for autism spectrum disorders: a randomized controlled trial. *J Am Acad Child Adolesc Psychiatry*. 2012;51:1052-65.

[27] Gharebaghy S, Rassafiani M, Cameron D. Effect of cognitive intervention on children with ADHD. *Phys Occup Ther Pediatr*. 2015;35:13-23.

[28] Cheung CH, Rijdsdijk F, McLoughlin G, Brandeis D, Banaschewski T, Asherson P, et al. Cognitive and neurophysiological markers of ADHD persistence and remission. *The British Journal of Psychiatry*. 2016;208:548-55.

[29] Wallander JL, Bann CM, Biasini FJ, Goudar SS, Pasha O, Chomba E, et al. Development of children at risk for adverse outcomes participating in early intervention in developing countries: a randomized controlled trial. *Journal of child psychology and psychiatry, and allied disciplines*. 2014;55:1251-9.

[30] Hintz SR, Kendrick DE, Vohr BR, Poole WK, Higgins RD. Gender differences in neurodevelopmental outcomes among extremely preterm, extremely-low-birthweight infants. *Acta paediatrica*. 2006;95:1239-48.

[31] Hwang LC, Bai CH, Chen CJ, Chien KL. Gender difference on the development of metabolic syndrome: a population-based study in Taiwan. *Eur J Epidemiol*. 2007;22:899-906.

[32] Bornstein MH, Han, C.H., & Haynes, O.M. Specific and general language performance across early childhood: Stability and gender considerations. *First Language*. 2004;24:267-304.

[33] Seller MJ, Perkins-Cole KJ. Sex difference in mouse embryonic development at neurulation. *Journal of reproduction and fertility*. 1987;79:159-61.

[34] Wright LL. Development of the sex difference in neuron numbers of the superior cervical ganglion: effects of transection of the cervical sympathetic trunk. *The Journal of comparative neurology*. 1987;263:259-64.

[35] Van Eden CG, Uylings HB, Van Pelt J. Sex-difference and left-right asymmetries in the prefrontal cortex during postnatal development in the rat. *Brain research*. 1984;314:146-53.

[36] Lei X, Smith JP, Sun X, Zhao Y. Gender Differences in Cognition in China and Reasons for Change over Time: Evidence from CHARLS. *Journal of the economics of ageing*. 2014;4:46-55.

[37] Ball RS. The Gesell Developmental Schedules: Arnold Gesell (1880-1961). *Journal of abnormal child*

---

psychology. 1977;5:233-9.

- [38] Ruess AL, Dally A, Lis EF. The Gesell Developmental Schedules and the physically handicapped child. *The American journal of occupational therapy : official publication of the American Occupational Therapy Association.* 1959;13:117-24 passim.
- [39] Tang D, Lee J, Muirhead L, Li TY, Qu L, Yu J, et al. Molecular and neurodevelopmental benefits to children of closure of a coal burning power plant in China. *PLoS One.* 2014;9:e91966.
- [40] Liu ZH, Li YR, Lu YL, Chen JK. Clinical research on intelligence seven needle therapy treated infants with brain damage syndrome. *Chin J Integr Med.* 2016;22:451-6.
- [41] Huo K, Zhang Z, Zhao D, Li H, Wang J, Wang X, et al. Risk factors for neurodevelopmental deficits in congenital hypothyroidism after early substitution treatment. *Endocrine journal.* 2011;58:355-61.
- [42] Portney LG, Watkins MP. *Foundations of clinical research : applications to practice:* Appleton & Lange; 1993.
- [43] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-74.
- [44] Lipsey MW. *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage; 1990.
- [45] Choe SA, Cho SI. Causes of child mortality (1 to 4 years of age) from 1983 to 2012 in the Republic of Korea: national vital data. *Journal of Preventive Medicine & Public Health.* 2014;47:336-42.
- [46] Paul R. *Language Disorders From Infancy Through Adolescence.* 3rd ed. St. Louis, MO: Mosby Elsevier; 2007.
- [47] Bayley N. *Bayley Scales of Infant and Toddler Development—Third Edition: Technical manual.* San Antonio, TX: Harcourt Assessment; 2006.
- [48] Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *J Pediatr.* 2012;160:553-8.
- [49] Houwen S, Visser L, van der Putten A, Vlaskamp C. The interrelationships between motor, cognitive, and language development in children with and without intellectual and developmental disabilities. *Research in developmental disabilities.* 2016;53-54:19-31.
- [50] Webster RI, Erdos C, Evans K, Majnemer A, Kehayia E, Thordardottir E, et al. The clinical spectrum of developmental language impairment in school-aged children: language, cognitive, and motor findings. *Pediatrics.* 2006;118:e1541-9.
- [51] Siegel LS. Infant perceptual, cognitive, and motor behaviours as predictors of subsequent cognitive and language development. *Canadian journal of psychology.* 1979;33:382-95.
- [52] Karasik LB, Adolph KE, Tamis-Lemonda CS, Bornstein MH. WEIRD walking: cross-cultural research on motor development. *Behav Brain Res.* 2010;33:95-6.
- [53] de Courten-Myers GM. The human cerebral cortex: gender differences in structure and function. *J Neuropathol Exp Neurol.* 1999;58:217-26.
- [54] Gaoua N, Herrera CP, Periard JD, El Massioui F, Racinais S. Effect of Passive Hyperthermia on Working Memory Resources during Simple and Complex Cognitive Tasks. *Frontiers Psychology.* 2017;8:2290-.
- [55] Chang KJ, Hong CH, Lee YH, Chung YK, Lim KY, Noh JS, et al. Memory Age Identity as a predictor of cognitive function in the elderly: A 2-year follow-up study. *Archives of gerontology and geriatrics.* 2018;74:68-71.
- [56] Frankenmolen NL, Fasotti L, Kessels RPC, Oosterman JM. The influence of cognitive reserve and age on the use of memory strategies. *Experimental aging research.* 2018;44:117-34.

- 
- [57] Zaehring J, Falquez R, Schubert AL, Nees F, Barnow S. Neural correlates of reappraisal considering working memory capacity and cognitive flexibility. *Brain imaging and behavior*. 2018.
- [58] Cragg L, Nation K. Language and the development of cognitive control. *Topics in cognitive science*. 2010;2:631-42.
- [59] Drover JR, Hoffman DR, Castaneda YS, Morale SE, Garfield S, Wheaton DH, et al. Cognitive function in 18-month-old term infants of the DIAMOND study: a randomized, controlled clinical trial with multiple dietary levels of docosahexaenoic acid. *Early human development*. 2011;87:223-30.