

# Is the Bayley Scales of Infant and Toddler Developmental Screening Test, Valid and Reliable for Persian Speaking Children?

Farin Soleimani,<sup>1</sup> Nadia Azari,<sup>1,\*</sup> Roshanak Vameghi,<sup>1</sup> Firoozeh Sajedi,<sup>1</sup> Soheila Shahshahani,<sup>1</sup> Hossein Karimi,<sup>1</sup> Adis Kraskian,<sup>2</sup> Amin Shahrokhi,<sup>1</sup> Robab Teymouri,<sup>1</sup> and Masoud Gharib<sup>3</sup>

<sup>1</sup>Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, IR Iran

<sup>2</sup>Department of Psychology, Karaj Branch, Islamic Azad University, Karaj, IR Iran

<sup>3</sup>Faculty of Paramedicine, Mazandaran University of Medical Sciences, Sari, IR Iran

\*Corresponding author: Nadia Azari, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, IR Iran. Tel/Fax: +98-2122180099, E-mail: farinir@yahoo.com

Received 2016 February 02; Revised 2016 March 12; Accepted 2016 April 27.

## Abstract

**Background:** Advances in perinatal and neonatal care have substantially improved the survival of at-risk infants over the past two decades.

**Objectives:** The purpose of this study was to assess the reliability and validity of the Bayley Scales of infant and toddler developmental Screening test in Persian-speaking children.

**Methods:** This was a cross-sectional prospective study of 403 children aged 1 - 42-months. The Bayley scales screening instrument, which consists of five domains (cognitive, receptive, and expressive communication and fine and gross motor items), was used to measure infants' and toddlers' development. The psychometric properties examined included the face and content validity of the scale, in addition to cultural and linguistic modifications to the scale and its test-retest and inter-rater reliability.

**Results:** An expert team changed some of the test items relating to cultural and linguistic issues. In almost all the age groups, cultural or linguistic changes were made to items in the communication domains. According to Cronbach's alpha for internal consistency, the reliability of the cognitive scale was  $r = 0.79$ , and the reliability of the receptive scale was  $r = 0.76$ . The reliability for expressive communication, fine motor, and gross motor scales was  $r = 0.81$ ,  $r = 0.80$ , and  $r = 0.81$ , respectively. The construct validity of the tests was confirmed using a factor analysis and comparison of the mean scores of the age groups. The intra- and inter-rater reliabilities of the Bayley Scales were good-to-excellent.

**Conclusions:** The results indicated that the Bayley Scales had a high level of reliability in the present study. Thus, the scale can be used in a Persian population.

**Keywords:** Screening Tools, Validity, Reliability, Infant

## 1. Background

Advances in perinatal and neonatal care have substantially improved the survival of at-risk infants over the past two decades (1). However, these advances have produced little change in the prevalence of developmental disorders among at-risk survivors (2, 3). Identifying infants at risk for developmental disabilities is the first step in providing services to maximize their physical and cognitive abilities and to minimize complications. Health conditions, such as a low birth weight, preterm birth, perinatal infection, and birth defects, increase the risk of 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 (4). In Iran, asphyxia, low birth weight,

preterm birth, and a high-risk pregnancy have been shown to adversely affect neurological development (5-7). The American Academy of Pediatrics recommends that pediatricians screen all infants and children during routine office visits for developmental problems (8). In the U.S., the emphasis has shifted to screening for disabilities at a younger age: from birth to 2 years (9).

Recent epidemiological data indicated that the rate of moderate-to-severe disabilities in at-risk infants in the early years of life was approximately 6.7% - 14% (10, 11). It has been estimated that more than 200 million children under 5 years do not reach their full potential in terms of growth, cognition, or socio-emotional development due to risk factors for neurological delay (12). In the U.S., about 13% of children aged 3 - 17 years were reported to have at least one developmental disability, and about 1.6% of children

were shown to have three or more developmental disabilities (13).

Robust estimates of the prevalence of development disabilities in less developed countries are rare. However, given the overall higher prevalence of most diseases of early childhood in less developed countries compared to developed countries, the rates are expected to be at least similar to, if not higher, than those in developed countries. The availability of adequate screening for developmental disabilities is limited in less developed countries where the expenditure on health is significantly lower than in developed countries. Research has shown that improved economic status has positive effects on child development in both developed and developing countries (14, 15) and that it may attenuate the negative effects of early developmental problems in the future (16). Given the higher rates of poverty in less developed countries, developmental disabilities may have substantial adverse effects on future health and socioeconomic outcomes. In addition to the paucity of data on pediatric neurological development, most extant data were collected using assessments developed for use in European or North American populations. Only a few psychometric tools have been developed specifically to measure neurological development in settings outside of Europe and North America.

There is a need for standardized, psychometrically sound developmental screening instruments that can be used by primary care providers for the early identification of infants with developmental problems in developing countries (17). In the present study, the Bayley scales was chosen as a screening instrument. The scales is an individually administered instrument, which assesses the cognitive, language, and motor functioning of infants and young children aged 1 - 42 months. It can be administered by a wide range of health professionals after limited training and in an acceptable time frame (18). The Bayley scales can be used to obtain detailed information about the functioning of children, even nonverbal infants.

In common with the majority of available psychometric tests, the Bayley scales originated in the Western world and was designed to suit the culture, language, and socioeconomic status of the respective populations. According to De Klerk (19), many tests can be adapted from one language and culture to another. However, individual scores based on tests supposedly measuring the same construct in various cultures cannot be interpreted at face value. The influence of culture on measuring specific psychological constructs needs to be explored to be able to adjust measurements to make them meaningful to the particular culture and to obtain equivalent or comparable measures across cultures. The two most important and fundamental characteristics of any measurement procedure are the re-

liability and validity of the scales. Any kind of assessment, whether traditional or "authentic," must be developed in a way that provides accurate information about the performance of the individual.

## 2. Objectives

This study was conducted for the purpose of cultural modifications and validation of the Bayley Scales for Persian-speaking children aged 1 - 42 months.

## 3. Methods

The Bayley screening Test is a subtest of the diagnostic Bayley scales of infant and toddler development (20). Items in the subtest have been shown to be particularly valuable in screening high-risk infants for developmental delay (18). The cut scores are used to determine whether the child shows competence in age-appropriate tasks, shows evidence of emerging age-appropriate skills, or shows evidence of being at risk for developmental delay. The infant's total score is then compared to norms in order to classify the child as competent or at risk of developmental delay. The test takes approximately 15 - 30 minutes to administer (15 - 20 minutes for children aged 12 months and younger and approximately 30 minutes for children aged 13 months and older) (18).

In this study, an expert team performed translation and back-translation, assessed the content and construct validity of the scale, and made cultural and lingual modifications. To assess the reliability, the internal consistency (Cronbach's alpha coefficient for each the five domains and each age group), test-retest, and inter-rater reliability were determined. A factor analysis and comparison of the mean scores of the groups were used to assess the validity. In accordance with other studies that used factor analysis and the Comrey sample size criterion (21), the sample size was determined to be 400 people in four age groups. The participants were selected from our centers using continuous sampling. A principal components analysis (PCA) was used to determine how many factors were significant in the test.

Prior to performing the factor analysis, the Kaiser-Meyer-Olkin measure of sampling adequacy was applied. The results yielded a sample adequacy value of 0.948 - 0.964. In addition, Bartlett's test of sphericity was significant ( $P < 0.0001$ ) in five domains. Therefore, a factor analysis of the correlation matrix of the questionnaire items was performed to ensure that none of the items was equal to zero. The Eigen-value and ratio of variance explained by each factor were used to determine the number of saturated significant factors in each section of the study test.

A pilot study was then carried out with 45 children aged 1 - 42 months to determine the degree of "clarity" of the items and their cultural appropriateness and to detect ambiguous items. The children were recruited using convenience sampling. The inclusion criteria were age 1 - 42 months of age, Persian speaking, and lacking developmental disorders. Informed consent was acquired from the parents of the children, and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki, as reflected by a priori approval by the institution's human research committee.

To test the study sample, a group of specialists, including experts in psychometrics and occupational therapists, was selected following an interview, and the group underwent a training course. The training course consisted of observations of training videotapes, observations of field assessments of 10 infants of various ages administered by an experienced psychologist and administration of each version to 10 infants of various ages. The raters were required to achieve agreement of > 90% compared with the results of the experienced psychologist.

After seeking the approval of health centers, the translated versions of the items were administered by the raters to the participants in the study.

#### 4. Results

Among the 403 children in the study, 195 (48.4%) were girls. Seventy-eight (19.4%) of the children were aged 7 - 12 months, and 125 (31%) were aged 1 - 6-months.

To determine the psychometric properties of the test, the items in each domain were translated to the Persian language and then back-translated by two independent native translators who also had experience in the field of child development. By comparing the two versions, the discrepant parts were identified and corrected. A panel of eight experts (two pediatricians, one psychologist, two speech pathologists, two pediatric occupational therapists, and one psychometrist) then assessed the content validity of the resulting Persian test. This expert team performed cultural and lingual modifications. Although the team attempted to preserve the headings in the original version, some modifications had to be made to ensure cultural compatibility and greater clarity of the Persian version. Most of the modifications in the domain "receptive and expressive communication" pertained to the language domain but were not limited to this domain. The modifications to the instructions of the receptive communication subscale were as follows: Unfamiliar games were replaced with more familiar ones, and the words "glass," "ball," "sweet," and "bird" were replaced with "cup," "cube," "cake", and "fish," respectively. These modifications were

in accordance with studies on vocabulary development in Persian-speaking children (22, 23). Another modification made to the instructions related to the expression of possession. Furthermore, as the Persian language has only one pronoun for both boys and girls, gender was mentioned, in addition to the pronoun. The modifications to the instructions of expressive communication subtests were changes made in auxiliary verbs (not usually used in Persian), continuous verbs in the administration manual, different signs for future and continuous present verbs in Persian grammar, and plurals because Persian-speaking children would have difficulty expressing these items. Some modifications were also made to the cognitive and motor domains, such as the replacement of traditional games and changes to pictures and storybooks.

The internal consistency of the Bayley subtests was assessed using the Cronbach alpha method. The reliability coefficients and standard error of measurement (SEM) are presented in Table 1. The stability of the scores of the Bayley Scales over time was assessed in a separate study of 45 children who were tested twice (4 - 7 days retest) by the same raters. The test-retest reliability was estimated using Pearson's correlation coefficient (Table 2). To determine the inter-rater reliability, two raters administered the revised version of the test to 36 children (Table 3).

In the PCA of the nature of the relationships between the test headings and to obtain definitions of the factors, it was assumed that coefficients greater than 0.3 were significant factors and that those less than 0.3 were random factors. Based on a range of 0.3 - 0.8 for the various headings of the five domains, we concluded that a single-factor model was the best structure for the Bayley Screening Test. In other words, this model was sufficiently reliably to assess the construct of the developmental progression of infants aged 1 - 42 months.

Table 4 presents the first Eigen-value and percentage of variance explained by the first factor in the PCA. The results indicated that a single-factor model was the best model in each domain for performing the factor analysis in each domain.

The nature and content of the test are concerned with progressive development. Thus, to determine the performance in the test according to chronological age, the scores of the age groups in the five domains were compared using a one-way variance analysis (ANOVA). The F index for cognitive, receptive, and expressive communication subtests and gross and fine motor subtests was 1202.74, 969.88, 826.61, 814.51, and 872.94 respectively ( $P < 0.01$ ). As the F index was larger than 0.01, with degrees of freedom of 3 and 399 (3.83), the null hypothesis was rejected with 99% confidence based on the equality in the mean scores of four age groups. The Scheffe post hoc test

**Table 1.** Reliability Coefficients and Standard Error of Measurement (SEM) of the Bayley Screening Subtests (n = 403)

Subtests	Alpha Coefficients	SEM	
		Boys	Girls
Cognitive	0.79	1.79	1.63
Receptive communication	0.76	1.40	1.49
Expressive communication	0.81	1.38	1.53
Fine motor	0.80	1.41	1.53
Gross motor	0.81	1.54	1.56

**Table 2.** Stability Coefficients of the Bayley Screening Subtests (n = 45)

Subtests	Pearson's Correlation Coefficient
Cognitive	0.98 <sup>a</sup>
Receptive communication	0.97 <sup>a</sup>
Expressive communication	0.97 <sup>a</sup>
Fine motor	0.97 <sup>a</sup>
Gross motor	0.98 <sup>a</sup>

<sup>a</sup>P < 0.01.

**Table 3.** Inter-Rater Coefficients of the Bayley Screening Subtests (n = 36)

Subtests	Pearson's Correlation Coefficient
Cognitive	0.993 <sup>a</sup>
Receptive communication	0.999 <sup>a</sup>
Expressive communication	0.991 <sup>a</sup>
Fine motor	0.998 <sup>a</sup>
Gross motor	0.990 <sup>a</sup>

<sup>a</sup>P < 0.01.

**Table 4.** Results of the Principal Components Analysis (PCA)

	Factors				
	Cognition	Receptive Communication	Expressive Communication	Fine Motor	Gross Motor
The first Eigen value	14.91	11.074	11.398	11.391	11.670
Percentage of explained variance	42.700	46.142	47.490	42.190	41.680

was used in a paired comparison of the mean test values. The differences between the mean values are shown in [Table 5](#).

## 5. Discussion

Screening tests are generally designed in two forms: objective and subjective. In objective tests, the examiners

directly observe and assess an infant's behavior. Example include the Denver II, early learning milestone scale for screening language skills, and Brigance, Battelle, and Bayley scales (24). Subjective tests are developmental questionnaires completed by parents, such as the parent evaluation of developmental status and the ages and stages questionnaire (ASQ). Parents' views of the developmental status of their infants have been considered appropriate

**Table 5.** Mean Test Comparison of the Scores of the Different Age Groups in the Five Domains

Age Group 1	Age Group 2	Mean Test Difference (Significance Level; P Value) <sup>a</sup>				
		Cognition	Receptive Communication	Expressive Communication	Fine Motor	Gross Motor
A	B	6.58	4.19	3.82	4.38	6.03
	C	14.27	10.87	10.56	10.67	12.16
	D	22.59	16.90	16.87	17.41	17.98
B	C	7.69	6.68	6.74	6.30	6.13
	D	16.01	12.71	13.05	13.03	11.96
C	D	8.32	6.03	6.30	6.73	5.28

<sup>a</sup>P < 0.01 for all values.

ate and reliable for years (25, 26). However, subjective tests have some weaknesses. For example, poorly educated parents may have difficulty reading a questionnaire, although this difficulty can be overcome by asking parents in an appropriate way. Furthermore, some physicians have opined that highly educated parents may be oversensitive to their infants' development and that the use of parent-based questionnaires can lead to increased referrals (25, 26). Thus, there are some doubts about the credibility of information provided by parents. It should be noted that questionnaires are used in two-stage screenings, and suspected or unsuccessful cases should be assessed through diagnostic tests or objective screenings that require greater amount of time and skill (25, 26).

There are no comprehensive tools applicable to all societies and all age groups, and there are no culturally compatible screening tools in many developing countries (25, 26). Thus, unstandardized tools should first be standardized according to the population of each country (25, 26). Developmental screening tools that have been translated into the Persian language include the Denver test and ASQ, whose criterion validity has not been verified. Denver II is an objective test for developmental screening of children from birth to the age of 8 years (27, 28). The sensitivity and specificity of this test have been reported to range from 40% - 83% to 40% - 80%, respectively (27, 28). In Iran, the psychometric properties of the Denver test were compared with those of the ASQ in a sample of 197 children (29, 30). The authors reported that the Kappa agreement between the two tests was poor (0.21), with agreement of 0.17 with the results of a physical examination. Therefore, the authors concluded that the Kappa agreement coefficient was poor in the Denver test. Due to its wide range of sensitivity and specificity, the Denver II test is not recommended (27).

The ASQ has been standardized in Iran. However, as a standard Iranian diagnostic developmental test is not available, the criterion validity, sensitivity, and specificity

of the Persian version have not been determined (31). The ASQ has some strong points. Unlike objective tests, it does not require the cooperation of the infant, and it has been designed according to developmental indices, which can be taught to parents. In addition, the ASQ is economical, and it can be administered in a short time. Its weaknesses include the need for a large space to store it, with its 4 - 5 pages (27). Furthermore, poorly educated parents may find it difficult to complete, and it is unable to detect developmental delay in 13% of children (27). Thus, its use in high-risk groups is uncertain.

The Bayley screening test items are a subtest of the cognitive, language, and motor items of the Bayley diagnostic test (18). In the U.S., the evidence of Bayley screening test validity was conducted to examine the relation between performance on the Bayley diagnostic and Bayley screening test. Scores of 1 - 4 in the Bayley diagnostic test were equivalent to the criterion used to define the at-risk category in the Bayley screening test, and Bayley diagnostic test scores of 5 - 7 were equivalent to the criterion used to define the emerging category (18). In that study, for children with Bayley diagnostic test scores of 1 - 4 (very low), the classification accuracy was moderate. The number of such children correctly identified by the Bayley screening test as being at risk ranged from 41.82% on the fine motor subtest to 65.91% on the receptive communication subtest, and none of these children was incorrectly classified as proficient. In the same study, for children with Bayley diagnostic test scores of 5 - 7, the Bayley screening test was even more accurate. The numbers of these children correctly identified as "emerging" ranged from 63.87 for the cognitive subtest to 77.78% for the receptive communication subtest. The numbers of such children misidentified as at risk was very low, ranging from 0.82% - 5.21%. For children with Bayley diagnostic test scores of 8 - 19, the Bayley screening test was very accurate, with 83.84% correctly identified as proficient in the cognitive subtest and 92.11% identified as proficient in

the receptive communication subtest (18). Furthermore, none of the children was incorrectly identified as at risk. Of note, in this classification, no child had a Bayley diagnostic test score of 1 - 4 (very low) and a Bayley screening test score in the component category, and no child had a Bayley diagnostic test scaled score of 8 - 19 (high) and a Bayley screening test score in the at risk category. This test was also shown to be valid in Taiwan, Canada, and the U.K. (32-34). In a study in the U.S., compared to the Alberta motor development scale, the Bayley motor subscale showed a higher correlation in early referral of high-risk infants to interventional service centers (34), confirming the suitability of its application in such cases.

An examination of the relation between a test's content and the construct it is intended to measure provides a major source of evidence for the validity of the test. Evidence of content validity is not based on empirical or statistics testing: rather, it is the degree to which the test items adequately represent and relate to the trait or function that is being measured. The test content also involves the wording and format of the items, as well as the procedures for administering and scoring the test. In the present study, the content validity of the test was confirmed by eight experts in child development, and the construct validity was confirmed using a factor analysis and comparison of the scores of the different age groups.

In the present study, in terms of the cultural and linguistic appropriateness of the items for Persian-speaking children, several items were modified. Other studies of screening tests, such as the ASQ, performed a similar process of item modification (35-37). According to the Scheffe post hoc test, there was a significant difference between the mean values, thereby indicating a correlation between the age and test scores in the five domains, with higher scores associated with increased age. These results confirmed the validity of the test construct. To confirm the reliability of the instrument, its internal consistency was determined, in addition to test-retest and inter-rater values. As shown by the assessment of internal consistency using Cronbach's alpha method, the reliability of the cognitive scale, receptive, and expressive communication scales was .79, .76, and .81, respectively, and the reliability of the fine motor and gross motor scales was .80 and .81, respectively, with a small SEM (< 2). A study in the U.S. reported similar reliability results, with good internal consistency (0.82 - 0.88) and test-retest reliability of 0.80 - 0.83 (18).

In this study, the inter-rater and intra-rater reliability coefficients of the test were excellent for all the subtests. The results indicate that raters who receive training in how to administer the Bayley scales can reliably assess Persian infants. The reliability data also suggest that the scores for the subtests reflect a high degree of internal consistency

in the items and that this version of the Bayley screening test is equally reliable for assessing individuals with different levels of development. In the present study, the Bayley screening test scores showed very good stability over time across the age groups. Thus, the results of the test provide a reliable measurement, and the scores a child obtains in the test can be interpreted with a high level of confidence.

One of the first cross-cultural psychometric studies of the application of the Bayley scales to infants in an Eastern setting was a study of term and preterm Taiwanese infants. In that study, the correlations between the BSID-II and Bayley-III raw scores were good-to-excellent for the cognitive and motor items and low-to-excellent for the language items. In addition, both intra- and inter-rater reliability showed good-to-excellent correlations (> 0.75) and small SEMs (< 2) for term and preterm Taiwanese infants aged 6 - 24 months (32).

The major strengths of the present study were the use of an objective assessment of five developmental domains and the inclusion of infants aged 1 - 42 months old children. The revised scale is appropriate for follow-ups of at-risk infants.

### 5.1. Conclusions

This study demonstrated high reliability, content validity, and construct validity for all the subtests of the Bayley screening test. The results indicate that the Bayley screening test is a reliable and valid tool for the assessment of child development in the Middle East.

### Acknowledgments

We acknowledge and appreciate the financial and executive support of the University of Social Welfare and Rehabilitation Sciences and the Pediatric neurorehabilitation research center. The present study relied on the work of many colleagues. These included Nobakht Z, Hajisadeghi A, Ghorbani N, Moradi M, and Nabati F. We also thank the health centers, caregivers, and families who cooperated and generously took the time to participate in this study.

### References

1. Chin JR, Swamy GK. Long-term survival and reproduction in preterm infants. 2009
2. Doyle LW, Roberts G, Anderson PJ, Victorian Infant Collaborative Study G. Outcomes at age 2 years of infants < 28 weeks' gestational age born in Victoria in 2005. *J Pediatr.* 2010;**156**(1):49-53. doi: [10.1016/j.jpeds.2009.07.013](https://doi.org/10.1016/j.jpeds.2009.07.013). [PubMed: [19783004](https://pubmed.ncbi.nlm.nih.gov/19783004/)].
3. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet.* 2008;**371**(9608):261-9. doi: [10.1016/S0140-6736\(08\)60136-1](https://doi.org/10.1016/S0140-6736(08)60136-1). [PubMed: [18207020](https://pubmed.ncbi.nlm.nih.gov/18207020/)].

4. Jelliffe-Pawlowski LL, Shaw GM, Nelson V, Harris JA. Risk of mental retardation among children born with birth defects. *Arch Pediatr Adolesc Med.* 2003;**157**(6):545-50. doi: [10.1001/archpedi.157.6.545](https://doi.org/10.1001/archpedi.157.6.545). [PubMed: [12796234](https://pubmed.ncbi.nlm.nih.gov/12796234/)].
5. Soleimani F, Vameghi R, Hemmati S, Hemmati S, Salman-Roghani R. Perinatal and neonatal risk factors for neurodevelopmental outcome in infants in Karaj. *Arch Iran Med.* 2009;**12**(2):135-9. [PubMed: [19249882](https://pubmed.ncbi.nlm.nih.gov/19249882/)].
6. Soleimani F, Vameghi R, Biglarian A. Antenatal and intrapartum risk factors for cerebral palsy in term and near-term newborns. *Arch Iran Med.* 2013;**16**(4):213-6. [PubMed: [23496363](https://pubmed.ncbi.nlm.nih.gov/23496363/)].
7. Soleimani F, Vameghi R, Biglarian A, Daneshmandan N. Risk factors associated with cerebral palsy in children born in eastern and northern districts of Tehran. *Iran Red Crescent Med J.* 2010;**20**(4):428-33.
8. Council on Children With D, Section on Developmental Behavioral P, Bright Futures Steering C, Medical Home Initiatives for Children With Special Needs Project Advisory C. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics.* 2006;**118**(1):405-20. doi: [10.1542/peds.2006-1231](https://doi.org/10.1542/peds.2006-1231). [PubMed: [16818591](https://pubmed.ncbi.nlm.nih.gov/16818591/)].
9. Yell ML, Shriner JG, Katsiyannis A. Individuals with disabilities education improvement act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus Except Child.* 2006;**39**(1):1-24.
10. Boulet SL, Schieve LA, Boyle CA. Birth weight and health and developmental outcomes in US children, 1997-2005. *Matern Child Health J.* 2011;**15**(7):836-44. doi: [10.1007/s10995-009-0538-2](https://doi.org/10.1007/s10995-009-0538-2). [PubMed: [19902344](https://pubmed.ncbi.nlm.nih.gov/19902344/)].
11. Larroque B, Ancel PY, Marret S, Marchand L, Andre M, Arnaud C, et al. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): a longitudinal cohort study. *Lancet.* 2008;**371**(9615):813-20. doi: [10.1016/S0140-6736\(08\)60380-3](https://doi.org/10.1016/S0140-6736(08)60380-3). [PubMed: [18328928](https://pubmed.ncbi.nlm.nih.gov/18328928/)].
12. Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B, et al. Developmental potential in the first 5 years for children in developing countries. *Lancet.* 2007;**369**(9555):60-70. doi: [10.1016/S0140-6736\(07\)60032-4](https://doi.org/10.1016/S0140-6736(07)60032-4). [PubMed: [17208643](https://pubmed.ncbi.nlm.nih.gov/17208643/)].
13. Boulet SL, Boyle CA, Schieve LA. Health care use and health and functional impact of developmental disabilities among US children, 1997-2005. *Arch Pediatr Adolesc Med.* 2009;**163**(1):19-26. doi: [10.1001/archpediatrics.2008.506](https://doi.org/10.1001/archpediatrics.2008.506). [PubMed: [19124699](https://pubmed.ncbi.nlm.nih.gov/19124699/)].
14. Currie J. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *J Econom Lit.* 2009;**47**(1):87-122.
15. Paxson C, Schady N. Cognitive development among young children in Ecuador the roles of wealth, health, and parenting. *J Human Resources.* 2007;**42**(1):49-84.
16. Feinstein L. Inequality in the early cognitive development of British children in the 1970 cohort. *Economica.* 2003;**70**(277):73-97.
17. McConnell SR. Assessment in Early Intervention and Early Childhood Special Education Building on the Past to Project Into Our Future. *Topics Early Child Special Educ.* 2000;**20**(1):43-8.
18. Bayley N, Reuner G. Bayley scales of infant and toddler development: Bayley-III. 7. San Antonio: Harcourt Assessment, Psych. Corporation; 2006.
19. De Klerk G. In: Online Readings in Testing and Assessment, International Test Commission. Born M, Foxcroft CD, Butter R, editors.; 2008. Cross-cultural testing.
20. Bayley N. Bayley scales of infant and toddler development In Technical Manual. San Antonio: Harcourt Assessment; 2006.
21. Comrey AL, Lee HB. A first course in factor analysis. Psychology Press; 2013.
22. Hayati L, Babazadeh M, Soleymanzadeh F, Farrokhi H. Sound growth-Persian speaking children aged 6 to 24 months in four sections. University of Social Welfare and rehabilitation Sciences; 1997.
23. Mehdipour N, Shirazi TS, Nematzadeh S. Most frequent expressing words of Farsi-speaking children ages between 18-24 months. *Speech Language Pathol Aut.* 2013;**1**(1):71-80.
24. eKolste K. Developmental Surveillance and Screening, Monitoring to Promote Optimal Development. University of Washington; 2004.
25. Drotar D, Stancin T, Dworkin P. Pediatric developmental screening: understanding and selecting screening instruments. Commonwealth Fund; 2008.
26. Oberklaid F, Efron D. Developmental delay-identification and management. *Aust Fam Physician.* 2005;**34**(9):739-42. [PubMed: [16184205](https://pubmed.ncbi.nlm.nih.gov/16184205/)].
27. Kliegman RM, Stanton BF, St Geme JW, Schor NF. Nelson Textbook of Pediatrics. Philadelphia: Elsevier; 2016.
28. Dworkin PH. 2003 C. Anderson Aldrich award lecture: enhancing developmental services in child health supervision—an idea whose time has truly arrived. *Pediatrics.* 2004;**114**(3):827-31. doi: [10.1542/peds.2004-0416](https://doi.org/10.1542/peds.2004-0416). [PubMed: [15342860](https://pubmed.ncbi.nlm.nih.gov/15342860/)].
29. Shahshahani S, Sajedi F, Azari N, Vameghi R, Kazemnejad A, Tonekaboni SH. Evaluating the Validity and Reliability of PDQ-II and Comparison with DDST-II for Two Step Developmental Screening. *Iran J Pediatr.* 2011;**12**(3):343-349.
30. Shahshahani S, Vameghi R, Azari N, Sajedi F, Kazemnejad A. Validity and Reliability Determination of Denver Developmental Screening Test-II in 0-6 Year-Olds in Tehran. *Iran J Pediatr.* 2010;**20**(3):313-22.
31. Sajedi F, Vameghi R, Habibollahi A, Lornejad H, Delavar B. Standardization and validation of the ASQ developmental disorders screening tool in children of Tehran city. *Tehran Univ Med Sci.* 2012;**70**(7).
32. 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**(11):3875-83. doi: [10.1016/j.ridd.2013.07.006](https://doi.org/10.1016/j.ridd.2013.07.006). [PubMed: [24029804](https://pubmed.ncbi.nlm.nih.gov/24029804/)].
33. 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**(4):553-8. doi: [10.1016/j.jpeds.2011.09.047](https://doi.org/10.1016/j.jpeds.2011.09.047). [PubMed: [22048046](https://pubmed.ncbi.nlm.nih.gov/22048046/)].
34. Jackson BJ, Needelman H, Roberts H, Willet S, McMorris C. Bayley Scales of Infant Development Screening Test-Gross Motor Subtest: efficacy in determining need for services. *Pediatr Phys Ther.* 2012;**24**(1):58-62. doi: [10.1097/PEP.0b013e31823d8ba0](https://doi.org/10.1097/PEP.0b013e31823d8ba0). [PubMed: [22207470](https://pubmed.ncbi.nlm.nih.gov/22207470/)].
35. Heo KH, Squires J, Yovanoff P. Cross-cultural adaptation of a pre-school screening instrument: comparison of Korean and US populations. *J Intellect Disabil Res.* 2008;**52**(Pt 3):195-206. doi: [10.1111/j.1365-2788.2007.01000.x](https://doi.org/10.1111/j.1365-2788.2007.01000.x). [PubMed: [18261019](https://pubmed.ncbi.nlm.nih.gov/18261019/)].
36. Kapci EG, Kucuker S, Uslu RI. How applicable are Ages and Stages Questionnaires for use with Turkish children?. *Topics Early Child Special Educ.* 2010;**30**(3):176-88.
37. Vameghi R, Sajedi F, Kraskian Mojembari A, Habiollahi A, Lornezhad HR, Delavar B. Cross-Cultural Adaptation, Validation and Standardization of Ages and Stages Questionnaire (ASQ) in Iranian Children. *Iran J Public Health.* 2013;**42**(5):522-8. [PubMed: [23802111](https://pubmed.ncbi.nlm.nih.gov/23802111/)].