ORIGINAL ARTICLE

Concurrent Validity of the Bayley Screening and the Bayley-III in Persian-Speaking Children

Farin Soleimani, MD [1]; Fatemeh Hassanati, PhD [1]; Zahra Ghorbanpour, PhD [2]; Mohsen Vahedi, PhD [3]; Nadia Azari, PhD [4]; Adis Kraskian, PhD [4]; Zahra Nobakht, PhD [1]

[1] Pediatrician, Research Professor, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

[2] Department of Biostatistics, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

[3] Department of Occupational Therapy, School of Rehabilitation Sciences, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

[4] Department of Psychology, Karaj Branch, Islamic Azad University, Karaj, Iran

Keywords:

Development

Neuropsychological tests

Child

Screening

Validity

Received:
05-Sep-2023

Accepted:
12- Feb-2024

Published:
11- Mar-2025

ABSTRACT

Objectives

This study aims to determine the concurrent validity of the Bayley-III and the Bayley screening in Persian-speaking children. Measuring child development with the Bayley-III can be time-consuming and expensive. Accordingly, this research seeks to assess the accuracy of the Bayley screening as a measure of developmental delay for high-risk infants by age group.

Materials & Methods

Concurrent validity between raw Bayley screening scores and Bayley-III scores was assessed by administering to 403 1-42 month-old children. The cut score of 1.0 standard deviation below the mean of two tests was calculated using the Bayley-III as the criterion measure. This study used 70% of the sensitivity and specificity cut scores to measure validity. In addition, the study calculated the Pearson and Cohen's kappa correlation for the association between the two measures.

Results

The sensitivity of cognitive, receptive, and expressive communication, fine, and gross motor scales of Bayley screening were 70.7, 81.4, 67.7, 60.7, and 58.1, respectively. Specificity values varied narrower, from 87.8% to 100%. The Cohen's kappa coefficient in all age groups was substantial. The Pearson correlation between two test scores is significant for all scales, although the coefficients are over 0.884. The sensitivity and specificity have no specific trends with children's age, and the best sensitivity concurrence on the two tests was in the 1-12 months old age range.

Conclusion

The study supports the concurrent validity of the Bayley screening, showing an assertive and significant association between Bayley screening and Bayley-III in Persian-speaking children.

How to cite this article: Soleimani F, Hassanati F, Ghorbanpour Z, Vahedi M, Azari N, Kraskian A, Nobakht Z. Concurrent Validity of the Bayley Screening and the Bayley-III in Persian-Speaking Children. Iran J Child Neurol. 2025; 19(2): 51-64. https://doi.org/10.22037/ijcn.v18i3.43167

Corresponding Author: Nobakht Z, PhD. Department of Occupational Therapy, School of Rehabilitation Sciences, Pediatric Neurorehabilitation Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. Email: farinir@yahoo.com

Introduction

Development in the first years of life has a long-term effect on health and quality of life (1). Developmental care is efficacious due to rapid brain development and brain flexibility in the first five years of life (2). Children with normal development have been shown to retain high levels of academic achievement, profession, and income (2-6), while children with developmental delays tend to be less successful later in life (4, 7-8). Inadequate nutrition, poor health status, and lack of parental interaction with children are factors systematically associated with developmental delay in developed and developing countries (2, 5, 9-10).

Previous research on developmental delays found that forty-three percent of children under-5 years of age worldwide are not growing to their full potential (11). Detecting developmental delays early is crucial, as the most effective interventions typically occur within the first few years of life (12). Selecting an effective and reliable method for measuring developmental delays is essential to achieve this. The usefulness of screening tools for diagnosing developmental disorders hinges on their ability to deliver accurate information (13). Preventive organizations consider evidence on the sensitivity and specificity of the screening instruments when considering their effectiveness in children's health promotion and early detection of developmental delays (14-15). Measuring childhood development with standardized tests is difficult and expensive. Seemingly, developmental screening is an easy and inexpensive early detection and intervention method. Thus, short tests or multi-dimensional screens are often used as alternatives. Several multi-dimensional tests evaluate children's development in different aspects, such as cognitive, language, and motor skills. The Bayley Scales of Infant and Toddler Development (Third Edition; Bayley-III) is the most widely used scale considering the gold standard (16-17).

Previous studies in Iran have shown that a large portion of children are affected by developmental delays. The prevalence rate of developmentally delayed 1-42 months children by Bayley-III using <-1 standard deviation were in cognition 19%, receptive communication 19.6%, expressive communication 21%, gross motor 19.2%, and fine motor 17.8% (18). The prevalence rate of developmentally delayed on <-0.67 SD below the mean by the Bayley screening test was on cognition 28.3%, receptive communication 20.2%, expressive communication 25%, gross motor 22.3%, and fine motor 26.9% (19). In another study, the prevalence of developmental delays in Iranian children screened by the Ages and Stages Questionnaire (ASQ) by means <-2SD ranged from 3.6% to 4.3% in different developmental domains (20).

Approximately 15 million infants are born preterm each year universally, accounting for approximately one in every ten babies born and the irresistible majority of high-risk infants. In 2016, the preterm birth rate in the United States was about 9.84%, which has a recent rise from 2014 to 2016 (21). Furthermore, the low birth weight (LBW) rate has increased over the past two decades because of an increased number of preterm births. Approximately 1.1 million neonates are born every year in Iran, and according to the available data, the frequency of LBW in Iran is 9% (22). Given the high rates in Iran, it may be necessary to implement more widespread screening tests for high-risk groups. If the Bayley screening can effectively evaluate the development of high-risk children, the policy implications could be significant. The Bayley-III and Bayley screening tests are the most commonly used tools for longitudinal follow-up of cognitive, language, and motor development for high-risk infants to evaluate early outcomes. Moreover, the ASQ may be used as a screening tool for children aged three months to five years to identify developmental delays (23). ASQ is a parent report questionnaire and has been successful in detecting cognitive and motor delays in the follow-up of premature and hypoxic-ischemic encephalopathy (24). However, it is less sensitive than quantitative tools such as the Bayley screener for definite high-risk populations to determine whether or not a more extensive assessment is needed (25).

Therefore, this study aimed to investigate the criterion/concurrent validity of the Bayley screening test as a screening measure for detecting developmental delay in high-risk Persian-speaking children. In this study, researchers administered the Bayley-III and Bayley screening tests to 403 children aged between 1 and 42 months. They then evaluated the validity of the Bayley screening test by comparing it to the Bayley-III, which was considered the gold standard. To understand the relationship between the two sets of results, they calculated Pearson and Cohen's kappa coefficients.

Materials & Methods

A representative sample of 403 children aged 1 to 42 months old was selected from healthcare centers in five geographic regions (North, South, West, East, and Center) of Tehran (the capital of Iran). The coverage of health visits in Tehran's health centers is 60%, and 98% of vaccinations are from birth to five years old. Sampling was relative to the children living in each region.

Children met eligibility criteria if they were apparently healthy and were Persian-speaking. Any child born without significant medical problems and not now diagnosed or cared for cognitive, motor, language, or behavioral problems was assessed to have normal development. Exclusion criteria were composed of developmental disorders such as attention deficit hyperactivity or autism spectrum disorders, chromosomal or congenital anomalies, intellectual disability, rehabilitation interventions, and sensory disorders such as hearing or vision complications. Information was obtained through medical record review.

Instruments

The Bayley-III and Bayley screening tests, which determined the adaptation, psychometric properties, and Persian-speaking cut scores in previous studies, were used (19, 26-28).

The Bayley-III is considered the gold standard for determining developmental delay in children aged 1 to 42 months on cognitive (91 items), receptive (49 items) and expressive (48 items) communication, fine (66 items), and gross motor (72 items) domains (14). The age range is divided into 17 age groups. Scores are obtainable as raw scores, scaled scores, composite scores, percentile ranks, age equivalents, and growth scores. Scaled scores are derived from the raw scores. The range of the scaled scores is 1 to 19, with a standard deviation (SD) of 3 and a mean value of 10. For children ages 12 months and younger, administration time ranges approximately 50 minutes, while testing time for children 13 months and older is 90 minutes. Materials for the test include a manual, stimulus, picture book, record forms, puzzle board, bell, cup, pegs, block series, disks, and some other toys for documentation of child development with play (17).

The Bayley screening has been applied as a standard tool in the developmental assessment of infants in pediatric populations, and it is specifically valuable in screening high-risk infants for developmental delay. This test determines if a child is ready for standard opportunities or if a more in-depth evaluation is necessary. The Bayley screening items are a subtest of the cognitive, language, and motor items of the Bayley-III. Item management and scoring criteria are the same across both instruments. This test consists of several subtests: 33 items on cognitive skills, 24 items on receptive communication, 24 items on expressive communication, 27 items on fine motor skills, and 28 items on gross motor skills (16). Two cut scores were determined for the nine age groups, separating scores into the three bands that determine the at-risk, emerging, and competent classes. Cut scores represent approximately 2 and 0.67 standard deviations below the mean. For children ages 12 months and younger, administration time ranges between 15 to 20 minutes, while testing time for children 13 months and older is approximately 30 minutes (16).

Procedure

Raters for this study were selected based on the following criteria: Occupational or speech therapy or psychology qualification, at least two years' work experience, and curiosity about working with children. All of the raters were native Persian speakers and had completed a one-week course of both theoretical and practical training. Experts supervised the practical training to evaluate whether raters could provide an optimum undertaking, observe management methods, and score the test. Based on the above elements, the experts determined whether the raters could administer the Bayley-III. Those who were unqualified did not operate for testing in our survey.

The test was accomplished in the healthcare centers near the child's living area. The Bayley-III reviews the child's age in days, prematurity, and performance on a series of tasks using standardized toy kits to establish raw scores. The primary caregiver was present to keep and console the child while trained raters conducted the Bayley-III. Raters obeyed all management necessities for both the Bayley screening and Bayley-III. In most cases, the sessions went according to the following plan: infants were assessed first with the Bayley screening and after two weeks by the Bayley-III.

In the early months of life, the rate of developmental changes is higher. Considering that the cut-off points of the Bayley screening have nine age groups from birth to 42 months, and the Bayley-III has 48 age groups, and the use of test items can be a learning aspect for the child, we chose two weeks between performing two tests.

Data Analysis

Means and standard deviations of the total raw score of Bayley-III and Bayley screening on cognitive, language (receptive and expressive communication), and motor (fine and gross) development were used to define the sample characteristics. This study used Cronbach's alfa to estimate internal consistency. It is scored between 0 and 1; scores closer to 1 indicate greater internal consistency, meaning more significant confidence in this scale.

Second, this research estimated delays individually of the two scales. In instructions for Bayley-III, the cut scores are lower than the mean by 1 SD and 2 SD. In the instructions for Bayley screening, the cut scores are lower than the mean by 2 and 0.67 SD. In the present study, both the Bayley-III and Bayley screening use lower than the mean by 1 SD criteria to determine delay. Although instructions for the Bayley screening did not set the score lower than the mean by 1 SD as a cut score, this study still uses it due to the following explanations. First, the 1-cut score equals the Bayley-III. Second, although 0.67 SD, in addition to 1 SD, was advised by the Bayley screening, 0.67 cuts are not generally used in the literature, while 1SD is more commonly used (16). Based on the above explanation, this investigation used the 1 SD cut instead of 0.67. Since Bayley-III and screening cuts scores specific to Persian-speaking children are available (27-28), the researchers used the criteria for defining delays (Bayley-III) and risk of delays (Bayley screening) in Persian-speaking children.

Third, this study measured the validity of the domains evaluated by the Bayley screening and Bayley-III on five scales: Cognitive, language (receptive and expressive), and motor (fine and gross) development. The cuts of 1.0 SD below the mean of two instruments for a measure of validity (sensitivity and specificity) were estimated using the Bayley-III as the criterion measure or gold standard. Sensitivity was estimated as the proportion of children with scores below the cuts on the Bayley screening raw score that also scored below the cuts on the Bayley-III raw score. Specificity was computed as the proportion of children with scores at or above the Bayley screening cuts who also had Bayley-III scores above the cuts. In the literature, no agreed-upon cuts are available for the rate of sensitivity and specificity that can be seen as acceptable to define the accuracy of detecting developmental delays. Evidently, few studies have set cut scores to determine the validity of psychometric tests. However, Steenis used 70% as the cut scores of sensitivity and specificity to estimate the validity of developmental screening scales (29-30). Following their methods, this study also arranged 70% as the cut score for deciding validity, meaning Bayley screening can be seen as a suitable screening instrument if the values for sensitivity and specificity are 70% or more.

In addition, this study computed the Pearson and Cohen's kappa correlation for the relationship between Bayley-III and Bayley screening tests. Regarding the effect size of the strength of the correlations, this research used the available guidelines: If the absolute value of correlations were between 0.1 - 0.3, 0.3 - 0.5, and above 0.5, then small/low, medium/moderate, and large/strong correlations exist retrospectively.

All statistical analysis was performed using Stata 14.1 statistical software. P-values below 0.05 were considered significant.

This study was approved by the Ethics Committee of the University of Social Welfare and Rehabilitation Sciences. Participants were assured that their personal information would remain confidential, and informed consent was obtained from all participants' parents or legal guardians.

Results

Table 1 summarizes the characteristics of the participants. As shown in Table 1, slightly over half (51.8%) were male. Mother's educational attainment was 47% moderate (13–16 years), 38% low (<12 years), and 15% high (17+ years).

The mean scores (SD) of the developmental outcomes of our sample are shown in Table 2.

Table 3 presents the internal consistency of the scales in Bayley screening (all four age groups)

Table 1. Characteristics of children in the study sample (n=403)

1. Toddler's age (4 age group)		N (%)
1-6 months (A)		125(31%)
7-12 months (B)		78(19.4%)
13-24 months (C)		111(27.5%)
25-42 months (D)		89(22.1%)
2. Boys, %		208(51.6)
	Low	153(38%)
3. mother Educational level [1]	Moderate	189 (47%)
	High	61(15%)

[1] Low educational level" refers to special education, primary school, or pre-vocational secondary education (< 12 years); medium educational level" refers to senior general secondary education, pre-university education, or secondary vocational education (13–16 years); high educational level refers to "higher professional education or university (17+ years)

Table 2. The summary statistics of Bayley screening and Bayley-III raw scores in four age groups (n=403)

Toddler's age	Cognitive		Receptive Language		Expressive Language		Fine Motor		Gross Motor	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Bayley Screening raw scores in 4 age group										
1-6 months (A)	5.59	2.72	4.92	2.42	4.09	2.15	5.64	2.56	6.35	3.07
7-12 months (B)	12.80	2.34	9.76	2.16	8.81	2.71	10.44	2.20	13.13	2.84
13-24 months (C)	20.51	3.15	16.49	3.19	15.46	3.24	16.94	2.95	19.27	2.56
25-42 months (D)	28.78	3.06	22.70	1.55	21.95	2.45	23.73	3.00	25.36	2.42
Bayley-III raw scores in 4 age group										
1-6 months (A)	15.33	8.078	6.72	2.626	4.96	2.281	11.63	7.674	14.95	7.595
7-12 months (B)	34.23	4.972	12.35	2.411	10.70	3.407	24.78	3.250	34.87	6.670
13-24 months (C)	52.35	7.435	21.63	5.060	22.20	6.235	34.51	4.132	51.69	5.393
25-42 months (D)	69.74	10.671	35.86	6.426	37.73	8.200	45.47	9.283	61.66	4.941

and Bayley-III (all age groups). As shown in all scales, Cronbach's alfa coefficient is higher than 0.96, indicating great internal consistency in the two instruments.

Table 4 presents the comparison risk of delay between the Bayley screening and Bayley-III under 1 SD and 2 SD cut scores.

The psychometric properties of the Bayley screening, including sensitivity and specificity, are presented in Table 5. Moreover, the study revealed

Table 3. Internal consistency of the Bayley Screening and Bayley-III (n=403)

scales	Cognitive	Receptive Language	Expressive Language	Fine Motor	Gross Motor
Bayley-III	0.98	0.97	0.98	0.97	0.98
Bayley Screening	0.97	0.96	0.96	0.96	0.96
Bayley Screening (4 age group)					
1-6 months	0.81	0.79	0.77	0.81	0.83
7-12 months	0.71	0.71	0.79	0.73	0.85
13-24 months	0.81	0.84	0.85	0.81	0.77
25-42 months	0.80	0.65	0.83	0.82	0.76

Table 4. Risk of delay between the Bayley Screening and Bayley-III in Persian language children (n=403)

Cutoff	Cognitive	Receptive Language	Expressive Language	Fine Motor	Gross Motor
< 2SD (%)					
Bayley-III	3.1	3.4	3.3	2.9	3.4
Bayley Screening	1.2	1.2	1.5	1.0	2.5
< 1SD (%)					
Bayley-III	14.0	14.8	15.1	13.9	13.3
Bayley Screening	14.1	15.9	12.9	15.1	13.6

Table 5. Pass/fail agreement between the Bayley Screening and Bayley-III (n=403)

Age	Scale	Sensitivity	Specificity	Cohen's kappa coefficient
1-42 months (n=403)	cognitive	70.7	95.4	0.665
	Receptive Language	81.4	95.3	0.741
	Expressive Language	67.7	97.1	0.694
	Fine Motor	60.7	92.2	0.510
	Gross Motor	68.1	98.2	0.730
Subgroups				
	cognitive	60.9	100	0.717
	Receptive Language	75.0	99.0	0.804
	Expressive Language	66.7	100	0.769

Continued Table 5.

1-6 months (n= 125)				
	Fine Motor	100	91.8	0.729
	Gross Motor	81.5	99.0	0.850
	cognitive	91.7	97.0	0.857
	Receptive Language	88.9	97.1	0.820
	Expressive Language	71.4	95.3	0.687
7- 12 months (n=78)				
	Fine Motor	76.9	93.8	0.687
	Gross Motor	46.7	100	0.586
	cognitive	75.0	95.8	0.708
	Receptive Language	81.2	94.7	0.722
	Expressive Language	69.2	93.9	0.592
13-24 months (n=111)				
	Fine Motor	30.4	98.9	0.386
	Gross Motor	66.7	99.0	0.739
	cognitive	57.1	87.8	0.308
	Receptive Language	90.0	89.9	0.612
25-42 months (n=89)	Expressive Language	64.3	98.7	0.712
	Fine Motor	40.0	94.5	0.126
	Gross Motor	66.7	94.6	0.629

that only three sensitivities fell below the 70% cut scores. These include expressive communication, as well as fine and gross motor skills, with scores of 67.7, 60.7, and 58.1, respectively. This study also obtained several findings regarding specific trends. Notably, Cohen's kappa coefficient was found to be significant across all age groups and scales. In age groups 13-24 and 25-42 months, Cohen's kappa coefficient on the cognitive and fine motor is moderate, and in the fine motor group, 25-42 months is low. Second, the receptive communication scale has a higher sensitivity than the other, and the gross motor scale has a higher specificity than the other, regardless of the cut score.

Next, this study examined the Pearson correlation between the Bayley screening and Bayley-III scores (Table 6). The correlations are significant for all domains, although the coefficients are over 0.884 for all scales.

Discussion

This study aimed to investigate whether the Bayley screening is an accurate screening measure for 1-42-month-old Persian-speaking children — by comparing results from the Bayley screening with child performance on the Bayley-III, a multi-dimensional test considered the "gold standard"

Table 6. Pearson Correlations between Bayley Screening and Bayley-III by scales (N=403)

Bayley Screening	Bayley-III				
	Cognitive	Receptive Language	Expressive Language	Fine Motor	Gross Motor
Cognitive	0.992**	0.950**	0.937**	0.939**	0.936**
Receptive Language	0.966**	0.965**	0.967**	0.893**	0.899**
Expressive Language	0.960**	0.961**	0.971**	0.882**	0.884**
Fine Motor	0.960**	0.954**	0.945**	0.946**	0.934**
Gross Motor	0.955**	0.954**	0.937**	0.953**	0.982**

* p<0.01,
**R UPU <0/001

for early childhood development. To fulfill this goal, we collected unique data on 403 children in Tehran city. Then, this study estimated the levels of development among the sample children using both Bayley screening and Bayley-III.

This study found that Bayley screening significantly correlates with Bayley-III in all scales in Persian-speaking children. In addition, the present study found that the sensitivity and specificity of the scales are above 70%, but only on expressive communication and motor scales, and the sensitivities are below 70% of the cut scores. In all age groups, Cohen's kappa coefficient is strong/large. In age groups 13-24 and 25-42 months, Cohen's kappa coefficient on the cognitive and fine motor is moderate, and in the fine motor group, 25-42 months was weak/low. The receptive communication scale has a higher sensitivity than the others, and the gross motor scale has a higher specificity.

Moreover, Sheldrick investigated the comparative accuracy of the ASQ-3, Parents' Evaluation of Developmental Status (PEDS), and Survey of Well-being of Young Children (SWYC): Milestones, on 1495 families of children aged nine months to 5.5 years. Among children under 42 months, the specificity of the ASQ-3 (89.4%) and SWYC Milestones (89.0%) was higher than that of the PEDS (79.6%), but differences in sensitivity were not statistically significant. Among 43-66 months children, the specificity of the ASQ-3 (92.1%) was higher than that of the SWYC Milestones (70.7%) and the PEDS (73.7%), but sensitivity to mild delays of the SWYC Milestones (54.8%) and of the PEDS (61.8%) was higher than that of the ASQ-3 (23.5%). Sensitivity exceeded 70% concerning severe (<70) delays, with 73.7% for the SWYC Milestones among younger children, 78.9% for the PEDS among younger children, and 77.8% for the PEDS among older children. The researchers found three frequently utilized screening questionnaires present good specificity but modest sensitivity for detecting developmental delays among children aged 9 months to 5 years (31).

In the current study, the sensitivity and specificity have no specific trends with children's age, and the best sensitivity concurrence on the two instruments was in the 1-12 months old range. These results show that the Bayley screening may be more accurate when used to test younger-aged children, though there is still a statistically significant correlation between the two tests in the older age. These results vary from studies that show that the sensitivity of screening tests increases with age (29, 32-34). The differences between our findings may be due to discrepancies between children's ages, sampling areas, or the criteria used to characterize delay. On the other hand, the obtained results are similar to those of Yue, who investigated the concurrent validity of the MacArthur communicative development inventory, the ASQ, and the Bayley-III and found that sensitivity is higher in younger (8-16 months old) than older (17-24 months) children (35).

This study's sensitivity values ranged from 30.4% to 100% on Bayley screening. Most subscales by age sensitivity values (17 out of 25) were above 70%. Furthermore, Cohen's kappa coefficient (22 out of 25) had a strong/large correlation, meaning that most children recognized as at risk for developmental delay according to the Bayley screening were identified as such by the Bayley-III (indicating high sensitivity).

Specificity ranged narrower, from 87.8% to 100%. Although all of the specificity values were greater than 70%, this indicates that all children who were not recognized as at risk for developmental delay by the Bayley screening did not show a delay on the Bayley-III (indicating high specificity).

Such ours, the classification accuracy of the Bayley screening with the Bayley-III on 1700 U.S. 1-42 months residents showed that, for children with Bayley-III scaled scores of 1-4 (very low), classification accuracy was moderate and ranged from 41.82% on fine motor subtest to 65.91% on receptive communication subtest and none of these children were misidentified as proficient. For children with Bayley-III scaled scores of 5-7, classification accuracy was even more accurate and ranged from 63.87% on the cognitive subtest to 77.78% on the receptive communication subtest. The percentage of children mistakenly identified as at risk was minimal, ranging from 0.82% to 5.21%. For those with Bayley-III scaled scores between 8 and 19, classification accuracy was notably high, ranging from 83.84% for the cognitive subtest to 92.11% for the receptive communication subtest. None of these children were incorrectly labeled as at risk (16).

This study shows that the Bayley screening test is capable of assessing the development of Persian-speaking children with great consistency or accuracy compared to the Bayley-III, with lower administration and cost-effectiveness. The Bayley screening is one of the best multi-dimensional screenings that can be used in high-risk children's surveys.

The strengths of the present study are its broad age sample (from birth to 42 months) and the fact that, for the first time, the validity of the Bayley screening experimented against the Bayley-III in Persian-speaking children. The second strength is that the Bayley-III and screening tests used in this study have been validated and normalized for a healthy reference population in Iran.

One limitation of this study was that it did not operate longitudinal follow-up data to experiment with the correlation between Bayley screening Test and Bayley-III.

In Conclusion

This study's findings suggest a strong and significant association between Bayley screening and Bayley-III in Persian language children. Specifically, Bayley screening performed well in assessing the development of children under 12 months old. This study also found that the sensitivity range (60.7-81.4) and specificity (92.2 - 98.2) were narrowly distributed. Thus, the Bayley screening properly assessed the children aged 1–42 months old who do not have a developmental delay according to the Bayley-III (high specificity), and it was able to accurately know many of those children who did have developmental delays according to the Bayley-III (high sensitivity). In recent years, the ASQs 2 and 3 have been used in healthcare centers in Iran for developmental screening. The ASQ is a parent report questionnaire; thus, it needs too little time and money and is suitable for mass developmental screening for the health system. The Bayley screener has good psychometric characteristics and is proposed for follow-up of high-risk infants.

Acknowledgment

We would like to thank the University of Social Welfare and Rehabilitation for financial support. Authors declare any sources of potential conflicting interest. The authors are appreciative to all children and parents who participated in this study. The University of Social Welfare and Rehabilitation Sciences Ethical Committee approved this study (N=801/93/18330/1).

Authors' Contribution

Soleimani F, Azari N, Kraskian A, and Nobakht Z, had substantial contributions to the conception and design of the work; they had substantial contributions to acquisition, analysis, interpretation of data and drafting the work and revising it critically for important intellectual content; they had also contributed for final approval of the version to be published; and they have agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Vahedi M, Hasanati F, Ghorbanpour Z had substantial contributions to the conception, design, drafting and revising the work; they had contributed for final approval of the version to be published; and they have agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

The authors declare that no conflict of or competing interests existed or occurred in conducting this study.

References

1. Almond D, Currie J. Human capital development before age five. Vol. 4b, Handbook of Labor Economics. 2011. 1315–1486 p.
2. Knudsen EI, Heckman JJ, Cameron JL, Shonkoff JP. Economic, neurobiological, and behavior perspectives on building America's future workforce. PNAS. 2006; 103(27):10155–62. https://doi.org/10. 1073/pnas.0600888103 PMID: 16801553
3. Cunha F, Heckman J, Schennach S. Estimating the technology of cognitive and noncognitive skill formation. Econometrica. 2010; 78(3):883–931. https://doi.org/10.3982/ECTA6551 PMID: 20563300
4. Kautz T, Lafontaine P, Raval D, Shaikh A, Smith J, Stigler S. Analyzing social experiments as implemented: A reexamination of the evidence from the High/Scope Perry Preschool Program. Quant Econom. 2010; 1(1):1–46. PMID: 23255883

5. Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020. “Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia.” American Economic Review 110 (1): 48–85.
6. Manning M., Patterson J. Lifetime Effects: The High/Scope Perry preschool study through age 40. Child Educ. 2006; 83(2):121. https://doi.org/10.1007/s11060-006-9265-3
7. Walker SP, Wachs TD, Grantham-Mcgregor S, Black MM, Nelson CA, Huffman SL, et al. Inequality in early childhood: Risk and protective factors for early child development. Lancet. 2011; 378 (9799):1325–38. https://doi.org/10.1016/S0140-6736(11)60555-2 PMID: 21944375
8. Gertler P, Heckman J, Pinto R, Zanolini A, Vermeersch C, Walker S, et al. Labor market returns to an early childhood stimulation intervention in Jamaica. Science. 2014; 344(6187):998–1001.
9. Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B. Developmental potential in the first 5 years for children in developing countries. Lancet. 2007; 369(9555):60–70. https://doi.org/ 10.1016/S0140-6736(07)60032-4 PMID: 17208643
10. Soleimani F, Vameghi R, Biglarian A, Daneshmandan N. Risk Factors Associated with Cerebral Palsy in Children Born in Eastern and Northern Districts of Tehran.. IRCMJ 2010; 12(4):428-433.
11. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, et al. Early childhood development coming of age: science through the life course. Lancet. 2017; 389(10064):77–90. https://doi. org/10.1016/S0140-6736(16)31389-7 PMID: 27717614
12. Heckman JJ. Policies to foster human capital. Res Econ. 2000; 54(1):3–56.
13. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making. 1991;11(2):88-94. doi: 10.1177/0272989X9101100203
14. US Preventive Services Task Force. 2015. Procedures manual, 1-84. https://www.uspreventiveservicestaskforce.org/Page/Name/procedure-manual. Accessed March 30, 2019.
15. Canadian Task Force on Preventive Health Care Procedure manual. https://canadiantaskforce.ca/wp-content/uploads/2016/12/procedural-manual-en_2014_Archived.pdf. Published March 2014. Accessed March 30, 2019.
16. Bayley N, Reuner G. Bayley scales of infant and toddler development, Screening Test; Technical Manual. San Antonio: Harcourt Assessment, Psych. Corporation; 2006.
17. Bayley N. Bayley Scales of Infant and Toddler Development: Technical Manual. The PsychCorp; 2006.
18. Soleimani, F., Azari, N., Kraskian, A., Karimi, H., Sajedi, F., Vameghi, R., Shahshahani, S., Mehdipour Shahrivar, N., Shahrokhi, A., Teymouri, R., Gharib, M., & Noroozi, M. (2022). A Comparison Study of the Tehran Norms to the Reference Norms on Children Performance of the Bayley III. Iranian Journal of Child Neurology, 16(2), 63-76. https://doi.org/10.22037/ijcn.v16i1.29236
19. Soleimani F, Azari N, Vameghi R, Sajedi F, Shahshahani S, Karimi H, Kraskian A, Shahrokhi A, Teymouri R, Gharib M. Are the Norms of Bayley-III Screening Test Appropriate for Persian Language Children? Iran J Child Neurol. Spring 2018; 12(2):91-

98).

20. F Sajedi, R Vameghi, A Kraskian Mujembari. Prevalence of undetected developmental delays in Iranian children - Child: care, health and development, 2014.
21. Brady JM, Barnes-Davis ME, Poindexter BB. The High-Risk Infant. Chapter 117, 3865-67. In Kliegman RM, ST Geme JW, Blum NJ, Shah SS, Tasker RC, Wilson KM. In: Nelson Text book of Pediatrics; 2020, Vol 2.
22. Iran-statistical-yearbook (Persian) 2020.
23. Bricker D, et al. Ages and Stages Questionnaires (ASQ): A parent completed, child monitoring system. 2nd ed. Baltimore: Paul H. Brooks; 1999.
24. Lindsay NM, et al. Use of the Ages and Stages Questionnaire to predict outcome after hypoxic-ischaemic encephalopathy in the neonate. J Paediatr Child Health. 2008;44:590) (Skellern CY, et al. A parent-completed developmental questionnaire: follow-up of ex-premature infants. J Paediatr Child Health. 2001;37:125.
25. Richard J. Martin, Avroy A. Fanaroff, Michelo C.Walsh. Early Childhood Neurodevelopmental Outcomes of High-Risk Neonates. In: Robert M.Kliegman. Fanaroff and Martin's Neonatal –Perinatal Medicine; 2015.p.1018-1032.
26. Azari. N, Soleimani F, Vameghi R, Sajedi F, Shahshahani S, Karimi H, Kraskian A, Shahrokhi A, Teymouri R, Gharib M. A Psychometric Study of the Bayley Scales of Infant and Toddler Development in Persian Language Children. Iran J Child Neurol. Winter 2017; 11(1):50-56.
27. Soleimani F, Azari N, Vameghi R, Barekati S H, Lornejad H, Kraskian A. Standardization of the Bayley Scales of Infant and Toddler Development for Persian Children. jrehab 2022; 23 (1) :8-31. URL: http://rehabilitationj.uswr.ac.ir/article-1-2890-en.html
28. Farin Soleimani, Nadia Azari, Roshanak Vameghi, Firouzeh Sajedi, Soheila Shahshahani, Hossein Karimi, Adis Kraskian, Amin Shahrokhi, Robab Teymouri, Masoud Gharib. Is the Bayley Scales of Infant and Toddler Developmental Screening Test, valid and reliable for Persian speaking children? Iran J Pediatr. 2016 October; 26(5):e5540:83-90.
29. Steenis LJP, Verhoeven M, Hessen DJ, van Baar AL. Parental and professional assessment of early child development: The ASQ-3 and the Bayley-III-NL. Early Hum Dev. 2015; 91(3):217–25. https://doi. org/10.1016/j.earlhumdev.2015.01.008 PMID: 25703316
30. Council on Children With Disabilities; Section on Developmental Behavioral Pediatrics; Bright Futures Steering Committee; Medical Home Initiatives for Children With Special Needs Project Advisory Committee . Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006;118(1):405-420. doi: 10.1542/peds.2006-1231
31. Sheldrick RC, Marakovitz S, Garfinkel D, Carter AS, Perrin EC. Comparative Accuracy of Developmental Screening Questionnaires. JAMA Pediatr. 2020;174(4):366–374. doi:10.1001/jamapediatrics.2019.6000
32. Rubio-Codina M, Araujo MC, Attanasio O, Muñoz P, Grantham-McGregor S. Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. PLoS One. 2016; 11(8):1–17.
33. Yue A, Jiang Q, Wang B, Abbey C, Medina

A, Shi Y, Rozelle S. Concurrent validity of the Ages and Stages Questionnaire and the Bayley Scales of Infant Development III in China. PLoS One. 2019 Sep 5;14(9):e0221675. doi: 10.1371/journal.pone.0221675. PMID: 31487302; PMCID: PMC6728026.

34. Hix-Small H, Marks K, Squires J, Nickel R. Impact of Implementing developmental screening at 12 and 24 months in a pediatric practice. Pediatrics. 2007; 120(2):381–9. https://doi.org/10.1542/peds.2006- 3583 PMID: 17671065

35. Yue, A, Luo, X, Jia, M, et al. Concurrent validity of the MacArthur communicative development inventory, the Ages and Stages Questionnaires and the Bayley Scales of Infant and Toddler Development: A study in rural China. Inf Child Dev. 2021; 30:e2219. https://doi.org/10.1002/icd.2219