Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

UDC 159.9.075 **doi**: 10.21702/rpj.2020.4.4

Original research article

A Psychometric Study of the Bayley Scales of Infant and Toddler Development – Third Edition in the Russian Federation

Polina A. Pavlova[1]*, Nadezhda I. Bakushkina[1], Ekaterina V. Suleymanova[1], Nadezhda V. Pavlova[1], Maria A. Lavrova[1], Inna V. Tuktareva[1], Dmitry A. Chegodaev[1], Olga A. Lvova[1,2], Dmitry M. Maksimov[2]
[1] Ural Federal University named after the first President of Russia B. N. Yeltsin, Yekaterinburg, Russian Federation
[2] Urals State Medical University of the Ministry of Healthcare of the Russian Federation, Yekaterinburg, Russian Federation
* Corresponding author. E-mail: polinaalexpavlova@gmail.com
ORCID ID: https://orcid.org/0000-0003-4263-0484, https://orcid.org/0000-0003-4578-6820, https://orcid.org/0000-0002-1718-1798, https://orcid.org/0000-0002-0577-3171, http://orcid.org/0000-0003-4781-9833, https://orcid.org/0000-0003-1984-517X, https://orcid.org/0000-0001-5123-127X, https://orcid.org/0000-0002-2280-3096, https://orcid.org/0000-0001-8525-2116

Abstract

Introduction. There is currently no universal comprehensive measurement tool for the assessment of children development in the Russian Federation (RF). The Bayley-III scales developed by American researchers are widely used as such a tool. Numerous research groups recognize the need to modify the original scales before using them in a new linguistic and socio-cultural environment.
Methods. The authors (a) translated the original Bayley-III manual into Russian, (b) tested the tool by assessing cognitive, language, and motor development of 163 Russian children aged 2–11 months, and (c) made an indirect comparison of the mean scale scores of neuropsychological development and those obtained from the original American sample of children using Student's t-test.
Results. The modified version of the Bayley-III manual has been successfully tested in the RF. The indices of language and motor development of the children examined in this study did not statistically differ from the original American data (10 points). Higher scores were obtained for the cognitive scale (10.7 versus 10; $p = 0.003$). However, this effect was not very pronounced (Cohen's $d = 0.25$).
Discussion. The indices of neuropsychological development of Russian children fully comply with the original Bayley-III norms, which opens up new possibilities for its use in the RF. Slightly higher scores of the cognitive scale among Russian children do not generally affect the compliance with the original tool, since the difference was not significant. The results of this study can be extrapolated to full-term Caucasian children aged 2–11 months, whose parents have at least secondary education and average level of earnings. The widespread use of Bayley-III requires its further adaptation in larger and more representative samples of children from different regions of the RF with the additional assessment of social-emotional development as well as adaptive behavior.

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

DEVELOPMENTAL PSYCHOLOGY

Keywords
Bayley-III scales, Bayley-3, neuropsychological development, infants, toddlers psychometric study, adaptation, developmental assessment, complex development, cognitive sphere, motor sphere

Highlights
- Examining the Bayley Scales of Infant and Toddler Development – Third Edition, the authors carried out the analysis of the structure of its scales, the original standardization procedure, and previous international experience of adaptation/psychometric studies of the tool.
- Testing the Russian version of Bayley-III involved the assessment of the neuropsychological development of 163 children aged 2–11 months living in Yekaterinburg and the Sverdlovsk region.
- The authors performed an indirect comparison of the data obtained in this study and the original Bayley-III norms.
- Cognitive, language, and motor scales scores demonstrated a high level of compliance with the Bayley-III norms and can be extrapolated to the population of children in the Russian Federation, taking into account the limitations.

For citation
Pavlova, P. A., Bakushkina, N. I., Suleymanova, E. V., Pavlova, N. V., Lavrova, M. A., Tuktareva, I. V., ... Maksimov, D. M. (2020). A psychometric study of the Bayley Scales of Infant and Toddler Development – Third Edition in the Russian Federation. *Russian Psychological Journal*, *17*(4), 49–64. doi: 10.21702/rpj.2020.4.4

Received: August 14, 2020
Revision received: October 23, 2020
Accepted: October 24, 2020

Introduction
Psychologists and the medical community have been long paying close attention to the initial period of child development. This interest is explained by the importance of timely diagnosis for developmental disorders in early childhood as well as by considerable genetic and phenotypic variability that determines the neurodevelopment of children. Although certain predictors of atypical development may be detected in infancy, many of these disorders are diagnosed quite late, in the second or third year of life. Besides, despite the presence of numerous methods assessing child development, there is no 'gold standard' for the complex assessment of child development in the Russian Federation (RF) (Kustova, Taranushenko, & Demyanova, 2018).

In view of the above considerations, the search for effective diagnostic tools capable of accurate identification of developmental disorders during the first years of life and their testing becomes highly important. Diagnosing developmental delay during this period, when the body's compensatory mechanisms are at a high level, can considerably increase the efficiency of corrective therapeutic interventions.

The Bayley Scales of Infant and Toddler Development – Third Edition, hereinafter Bayley-III (Bayley, 2006), is the most studied diagnostic tool used all over the world. Bayley-III is widely recognized as a 'gold standard' for assessing the general development of children aged 16 days to 3.5 years (Azari et al., 2017; Ranjitkar et al., 2018; Yue et al., 2019). This measurement tool is based on well-known and generally accepted developmental theories (D. Bruner, L. S. Vygotsky,

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

A. R. Luria, J. Piaget) and is consistent with the results of research in child development, including studies in neuropsychology and in information processing (Aylward, 1988; Colombo & Cheatham, 2006; Colombo & Mitchell, 2009), functional socio-emotional theory (Greenspan, DeGangi, & Wieder, 2001), and the theory of adaptive behavior (Weiss, Oakland, & Aylward, 2010).

A significant limitation for the widespread application of the Bayley-III in the RF is the lack of age-specific norms for the Russian population of children. Cross-cultural studies show the need to adapt this assessment instrument, taking into account language differences and social characteristics of the population (Azari et al., 2017; Hegde, Rao, Raguram, & Gangadhar, 2013; Hoskens, Klingels, & Smits-Engelsman, 2018; Fuiko et al., 2019; Steenis, Verhoeven, Hessen, & van Baar, 2015; Sun et al., 2019). Application of the Bayley-III without prior standardization can lead to misinterpretation of results in the form of overestimation or, conversely, underestimation of a child's level of development.

Despite the fact that the Bayley-III has been actively used in the world since 2006, its proper adaptation has not yet been carried out in the RF. However, the Bayley-III is actively used for research and clinical purposes (no. BN40703). In view of this limitation, most Russian authors use the Bayley-III as an experimental tool rather than a diagnostic instrument, using its raw scores for statistical analysis (Kiselev et al., 2016; Belousova & Shvets, 2019; Bakushkina, Kiselev, Lvova, Suleimanova, & Tuktareva, 2018; Kiselev, Lvova, & Bakushkina, 2016). Several papers presented the converted raw scores, which ranges were identified on the basis of the American normative data (Zavadenko, Medvedev, & Degtyareva, 2018; Kosyakova & Bespalova, 2019; Shifman, 2016).

Thus, the importance of the Bayley-III application and present limitations of its practical employment as a diagnostic tool dictate the need for the appropriate adaptation and psychometric studying procedures in the Russian population of children. As the first stage of this procedure, the authors of this study analyzed the structure of the Bayley-III scales, examined the original standardization procedure, and conducted a systematic analysis of the experience of its adaptation and psychometric study in previous international studies. Prior to the application of the assessment scales in the Russian sample, the authors translated the original Bayley-III manual and evaluation forms into Russian. As a final stage, the authors carried out an experimental study of the neuropsychological development of 163 children from 2 to 11 months using the modified Bayley-III; the findings were compared with the original indices obtained from the sample of American children.

The Bayley-III design

The Bayley-III utilizes an efficient administration design that relies on age-based starting points, reverse principle, and discontinue criteria (Bayley, 2006).

The Bayley-III includes five scales. Each scale contains a certain number of items. Thus, the Bayley-III includes the following scales: (a) cognitive scale (91 items); (b) language scale (with subscales of receptive language (49 items) and expressive language (48 items)); (c) motor scale (with subscales of fine motor (66 items) and gross motor (72 items)); (d) social-emotional scale (35 items); and (e) adaptive behavior scale (241 items). Cognitive, language, and motor scales directly assess a child's performance on items, which makes it possible to evaluate the level of neuropsychological development (NPD). The social-emotional and adaptive behavior scales use indirect assessment presented in the form of a questionnaire filled out by a specialist based on caregivers' answers about the characteristics of their child's behavior in everyday life.

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

If the child performs the items correctly, one point is scored. The points for each scale are summed up. The so-called 'raw' scores are converted into (a) scaled scores that establish the boundaries of normative development; (b) composite scores used to compare all the scales and to identify individual characteristics of a child's development; (c) percentiles that make it possible to estimate the frequency of occurrence of the indices obtained by a child in the population corresponding to the standardizing sample.

Standardization of the Bayley-III in the USA

The recruitment of participants in the original standardization of the Bayley-III was carried out in the United States of America (USA) between January and October 2004 based on 2000 US census data (Bayley, 2006). The sample was stratified by age, gender, ethnicity, geographic region, parental or caregivers' education level. The sample consisted of 17 age groups with letter designations from A to Q, including 1,700 children aged 16 days to 43 months and 15 days. In each group, children were gender-equal. Most of the participants were Caucasian; their parents had average or above average levels of education and earnings.

The goals and objectives of the subsequent application of the measurement tool, – namely, the need to assess the compliance of a child's development with age norms, explained the fact that the majority of participants were 'typically developing' children. This category included children born at 37–42 weeks of gestation, without considerable neurological and somatic pathology, who did not receive treatment for mental, physical or behavioral disorders, and did not have the following diagnoses: attention deficit hyperactivity disorder (ADHD); chromosomal abnormalities; diseases caused by prenatal exposure to toxic substances (including fetal alcohol syndrome); congenital malformations of the central nervous system (CNS); genetic or congenital diseases; mental retardation; intraventricular hemorrhage; pathologies of the respiratory system; severe attachment disorders; severe sensory impairment; low birth weight; prematurity.

Children were excluded from the main research sample if they received assistance under the Early Childhood Intervention program, had risk factors affecting the development of the central nervous system, and if they were taking medications that could affect the tolerance of physical and intellectual activity, received inpatient treatment, had hearing or vision impairments, and their parents were not native English speakers.

Approximately 10 % of the sample consisted of children with clinical diagnoses, including Down syndrome, cerebral palsy, autism, Asperger syndrome, Rett syndrome, cognitive epileptiform disintegration, prematurity, fetal alcohol syndrome, specific speech disorders, asphyxia, hypotrophy, etc. We enlarged our experimental sample by including children of this category in order to increase its representativeness.

Psychometric and Adaptation studies of the Bayley-III

Because of its high diagnostic value the Bayley-III is widely used throughout the world. Table 1 presents previous studies attempted to adapt or psychometrically test the Bayley-III in other countries.

Table 1
The Bayley-III adaptation/psychometric publications

Authors, country, year	Sample characteristics	Procedure and results
Hua et al., China, 2019	N = 1444; age groups are formed according to the Bayley-III manual. *Inclusion criteria:* full-term children; absence of medical complications and clinical diagnosis at the time of examination, absence of treatment for mental, physical, or behavioral difficulties. *Exclusion criteria:* comorbidities or risk factors for development, such as hearing or visual impairment, medication that affects behavior, hospitalization or infection during testing, nutritional problems, sleep problems.	The authors made both forward and backward translation and cultural adaptation. Children were assessed only with the cognitive scale. Test-retest reliability, correlation and reliability of criteria for compliance with Bayley-III psychometric criteria were performed in 5–10 % of children randomly selected from the entire sample. *Conclusions:* The Bayley-III cognitive scale can be used to assess development as well as to account for gender differences when used in a Chinese sample.
Ranjitkar et al., Nepal, 2018	N = 600; 6–11 months of age. *Inclusion criteria:* age up to 1 year; residence in Bhaktapur municipality and surrounding areas for the next 12 months; signed informed consent. *Exclusion criteria:* severe systemic disease requiring hospitalization; severe malnutrition; taking vitamin B12 supplements; severe anemia; ongoing acute infections requiring medical treatment.	The authors made both forward and backward translation and adaptation according to cultural characteristics. The sample comprised 10 % of premature babies. *Conclusions:* the scores on the cognitive and motor scales are comparable to American norms. The indices on the language scales were significantly lower than those in the American sample and require special attention in interpretation. In general, Bayley-III can be used on Nepalese children 6–11 months of age. Cultural adaptation and standardization of the scales are prerequisites for a more valid and reliable assessment using this tool.

Table 1
The Bayley-III adaptation/psychometric publications

Authors, country, year	Sample characteristics	Procedure and results
Krogh et al., Denmark, 2012	N = 45; longitudinal study: 4, 7, 10 and 13 months of age. *Exclusion criteria:* prematurity; the presence of physical/mental disorders.	The authors made only forward translation. 55.1 % of parents in the Danish sample had 16 or more years of getting education. In the original American sample this percentage was 27.6 %. *Conclusions:* significant differences were obtained between Danish and American norms on all scales. In particular, Danish children have lower scores on the receptive language scale for all the ages. Bayley-III should be used with caution in non-US samples.
Chinta et al, Australia, 2014	N = 156; 3 years of age. *Inclusion criteria:* full-term healthy children, without chromosomal abnormalities, having no history of surgical interventions.	No adaptation was carried out, the original version and American norms were used to assess the development of Australian children. *Conclusions:* Australian children obtained higher scores on all the scales, except for the motor scale. Applying American norms can lead to an underestimation of mild disorders. Adaptation including standardization is required for local samples.
Azari et al., Iran, 2017	N = 403; age from 1 to 42 months (all the A–Q age groups of the original Bayley-III). The least number of children (N = 10) is in the age group E. The maximum number of children (N = 38) is in the age group D. *Inclusion criteria:* 1 to 42 months of age; normal development and the absence of any obvious developmental disorders; native Persian speakers.	The authors made both forward and backward translation, and adaptation according to cultural features. *Conclusions:* The Bayley-III is a valid and reliable tool for assessing the development of Persian-speaking children.

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

Table 1
The Bayley-III adaptation/psychometric publications

Authors, country, year	Sample characteristics	Procedure and results
Yu et al., Taiwan, 2013	N = 178 (preterm), N = 62 (term-born). Longitudinal study: 6, 12, 18, and 24 months (with age-adjustment for the preterm infants). *Inclusion criteria (full-term):* birth weight over 2500 g; gestation period 38–42 weeks; absence of serious prenatal and perinatal complications. *Inclusion criteria (preterm):* birth weight less than 1500 g; gestational period less than 37 weeks; joining the study within 7 days of birth; a child in a single pregnancy or the first child in a multiple pregnancy; no congenital anomalies or severe neonatal diseases. *Additional criteria for all the participants:* mothers over 18 years of age who can read and speak Chinese; no past history of alcoholism and/or drug addiction; parents who are married at the time of a child' birth.	Assessment done with the Bayley-III, and additionally with the second version of the tool. Thus, the results of both evaluations were compared and a conclusion about the potential for application of each of them was made. *Conclusions:* The Bayley-III is a reliable tool for assessing the development of full-term and premature Taiwanese children aged 6–24 months.
Steenis et al., Netherlands, 2015	N = 1912; age from 14 days to 42 months 14 days; 17 age groups according to the Bayley-III manual. *Inclusion criteria:* according to parents' report, a child has no problems with physical or mental health, has no regular medication intake; birth weight above 2500 g; gestational age – at least 37 weeks.	The analysis was carried out using Dutch and American norms. The Dutch sample as well as the American one comprised 10 % of children at risk or with developmental delays. *Conclusions:* Dutch norms differ from American ones by all the 5 scales; the differences are clinically significant. The importance of population-specific norms for the interpretation of developmental assessment results was shown.
Ballot et al., South Africa, 2017	N = 74; each child was assessed at least once. The evaluation of the development was carried out at the age of 9–12 and 15–20 months. *Inclusion criteria:* full-term infants discharged within 48 hours of birth. *Exclusion criteria:* obvious developmental anomalies.	*Conclusions:* Bayley-III is a reliable tool for assessing the development of South African urban children.

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

Table 1
The Bayley-III adaptation/psychometric publications

Authors, country, year	Sample characteristics	Procedure and results
Hanlon et al., Ethiopia, 2016	N = 896; 2 age groups: 30 months (N = 440), 42 months (N = 456). Inclusion criteria: normal weight and height at birth.	The authors made both forward and backward translation and modified stimulus materials (for example, replaced the images in a picture book) in accordance with cultural characteristics. *Conclusions:* the adjusted version of the Scales has shown its reliability and applicability for Ethiopian children.
Sun et al., Vietnam, 2019	N = 267; age 3–43 months. *Inclusion criteria:* children under 4 years of age from District 8 of Ho Chi Minh City. *Exclusion criteria:* history of chronic severe illness (e.g., congenital heart disease, epilepsy); prematurity; previous hospitalization in intensive care; developmental delay.	Adaptation through forward and backward translation; modifications with regard to cultural features, results of the pilot study. The structure of the adapted instrument differs from the original one. However, the differences are consistent with the theories of early development. *Conclusions:* The adapted version of Bayley-III shows high reliability, and also meets strict criteria of invariance for gender and age groups.

Of all the variety, there is only one study that performed adaptation close to the original procedure, including forward and backward translation, cultural adaptation, the analysis of all the ages according to the original tool, as well as 10 % of children with developmental disabilities included into a sample (Steenis et al., 2015); the evaluation at all age ranges corresponding to the original Bayley-III scales has been done in three studies (Azari et al., 2017; Hua et al., 2019; Steenis et al., 2019); forward and backward translations were performed in four studies (Azari et al., 2017; Hua et al., 2019; Steenis et al., 2015; Sun et al., 2019); two studies used longitudinal design (Krogh, Væver, Harder, & Køppe, 2012; Yu et al., 2013); in one study, the maximum number of children in age groups was 38 (Azari et al., 2017). Only one study comprised 10 % of children with clinical diagnoses (Steenis et al., 2015); another study comprised 10 % of premature children (Hua et al., 2019).

The findings of the analysis of the Bayley-III adaptation or psychometric studies indicate that in most cases the scores of the assessed children differed from the American norms on either all the scales or the majority of them (Chinta, Walker, Halliday, Loughran-Fowlds, & Badawi, 2014; Krogh et al., 2012; Steenis et al., 2015). On the other hand, many authors have demonstrated the possibility of successful application of American norms on a local sample without prior cultural adaptation (Ballot et al., 2017; Yu et al., 2013). In general, most of the studies indicate the reliability of the instrument after its adaptation (Azari et al., 2017; Hanlon et al., 2016; Hua et al., 2019;

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

Ranjitkar et al., 2018; Sun et al., 2019). Thus, most of the Bayley-III adaptations are incomplete and, despite the conclusions about the possibility of their use, the diagnostic results should be interpreted with caution.

Methods

The data collection was carried out within the Longitudinal Study of Neurocognitive Development in Children at the Laboratory of Brain and Neurocognitive Development, Department of Psychology, Ural Federal University named after the first President of Russia B. N. Yeltsin (UrFU). The recruitment of children was carried out using the resources of the Department of Psychology of the Ural Federal University, the Department of Pediatric Neurology and Neonatology of the Urals State Medical University, and the Yekaterinburg Clinical Perinatal Center (Children's City Hospital no. 10). The project was approved by the ethics committee at the Ural State Medical Academy of the Ministry of Health and Social Development of the Russian Federation. We provided the legal representatives of children participating in the study with information about the goals and methods of the project and explained them the principle of maintaining anonymity. We obtained a voluntary informed consent from a parent or a caregiver of a child before their participation in the project.

Current analysis was carried out on 3 scales (with all the corresponding subscales) – (a) cognitive, (b) language, and (c) motor. The scores of the scales based on the parent/guardian reports (social-emotional and adaptive behavior scales) were not used due to their lower reliability. The evaluation with the assessment tool was conducted by the researchers of the laboratory who received certified training in the use of Bayley-III as a psychodiagnostic tool. The study was carried out in a specially equipped room with parallel video recording of the process for the subsequent additional analysis of behaviors of children and parents.

The Bayley-III psychodiagnostic tool used in the study was officially purchased from a publishing house (England); its diagnostic materials and the content of single items was appropriate for the Russian socio-cultural environment. Thus, we used the original stimulus materials corresponding to the administration rules of the Bayley-III with evaluation forms translated into Russian.

The forward translation of the Bayley-III manual and its evaluation forms was carried out with preservation of contextual meaning. Difficulties in the modification of the tool were primarily associated with the grammatical differences in English and Russian languages. For example, in item 34, the subscale of receptive language, and item 34, the subscale of expressive language, the original version assumes an assessment of a child's understanding and use of the present continuous, which has no analogue in Russian speech. In Russian, the present tense is preferable when describing static situations and events occurring at the moment, which is used when carrying out diagnostics. In this regard, after the translation, these items were modified as follows: in the first case, we analyzed a child's understanding of verbs in the present tense and in the second case – their use. These items are intended for children over 1.5 years of age.

For the recruitment of study participants, we applied following inclusion criteria: full-term babies (beyond 37 weeks of gestation, according to the WHO criteria) from 2 months 16 days to 10 months 30 days, whose anthropometric parameters at birth correspond to the physiological norm.

To increase the representativeness and compliance with the original Bayley-III, 10 % of children with a family risk of ADHD or autism spectrum disorders (ASD), as well as with clinical diagnoses of ischemic stroke and prematurity were included into the sample.

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

The criteria for inclusion in the project corresponded to the pediatric norms that meet the WHO recommendations in order to compare the study results with the original data (Bayley, 2006; Volodin, 2009).

The recruitment of children was carried out by the pediatricians, neurologists, neonatologists from medical institutions in Yekaterinburg. The primary documentation of the study was a survey containing information on anthropometric and clinical characteristics of children at birth, anamnesis, information about the presence of ASD and ADHD were filled in.

In terms of the main socio-economic and demographic characteristics of the families, the scores obtained in our sample of children were mostly comparable with the American data (Bayley, 2006; Byers-Heinlein & Lew-Williams, 2013; Rao, Hammen, & Poland, 2009; Savostyanov et al., 2018). Nevertheless, our sample had certain differences. Thus, the parents of the children participating in our study were the educated urban population of Yekaterinburg and the Sverdlovsk region, while the original tool involved representatives of all social strata and regions of the United States. Besides, in the American sample a wider range of diseases, including more severe disorders, was presented among children with neurological pathology.

In general, our sample comprised 163 Caucasian children of 2–11 months, who lived mainly in Yekaterinburg and whose parents or caregivers have secondary or higher education and average or high level of earnings.

Statistical analysis was carried out using the Gretl 2019d software (Allin & Riccardo, 2020). The Student's t-test was used for indirect comparison of the mean scale scores with the American standard scores (which were 10 (± 3) points for all the scales in all age groups). The statistical significance of differences was established at $p < 0.05$ level. When calculating differences for the scale scores, we used estimates of effect size (Cohen, 1988).

Results

The final sample comprised 163 infants in the age from 2 months 24 days to 10 months 29 days (64 % of boys), of whom 10 % had a family risk of ASD and ADHD or a clinical diagnosis. Table 2 presents anthropometric and clinical characteristics of the children at birth. The parameters of normally developing children at birth met the WHO criteria for full-term pregnancies. Children with the pathology had either a clinical diagnosis (ischemic stroke, n = 4; prematurity, n = 5) or disease-specific family history (risk for ASD, n = 4; risk for ADHD, n = 2). Compared to 'normal' children, Apgar scores for children with a clinical diagnosis and disease-specific family history were higher by 1 point.

Table 2
Anthropometric and clinical characteristics of children at the time of birth

Index	Mean (SD); range
Gestational age, weeks	39 (1); 34–42
Body weight, g	3280 (440); 1830–4228
Body length, cm	51 (2); 42–58

Table 2
Anthropometric and clinical characteristics of children at the time of birth

Index	Mean (SD); range
1-minute Apgar score	6 (1); 5–8
5-minute Apgar score	8 (1); 5–9

Note: SD refers to standard deviation.

The NPD characteristics of Russian children generally corresponded to the original data on the three main scales – cognitive, language, and motor (Table 3). However, for Russian children the mean score of cognitive development was slightly higher than the original American norms in the corresponding age range (10.7 versus 10), which was confirmed statistically ($p = 0.003$). However, in terms of the effect size, this difference was poorly pronounced (0.7 points; Cohen's $d = 0.25$). There were no statistically significant differences in the subscales of receptive and expressive language, and fine and gross motor subscales (Table 3).

Table 3
Comparison of the Bayley-III scale scores between Russian and American samples

Scales	Study scale scores (1) (n = 163)	Original scale scores (1) (n = 600)	*p* (2)
Cognitive	10.7 (2.57) CI 95 % [10.3–11.1]	10 (3)	0.003
Receptive language	9.67 (2.24) CI 95 % [9.32–10]	10 (3)	0.1
Expressive language	10.2 (1.95) CI 95 % [9.85–10.5]	10 (3)	0.3
Fine motor	10 (2.24) CI 95 % [9.69–10.4]	10 (3)	1
Gross motor	9.98 (2.57) CI 95 % [9.58–10.4]	10 (3)	0.9

Note: (1) group mean (GM), confidence interval (CI) – only for the RF; (2) statistical significance differences, Student's t-test.

Russian Psychological Journal, 2020, Vol. 17, No. 4, 49–64. **doi**: 10.21702/rpj.2020.4.4

Discussion

Within current research we conducted the psychometric study of the scales of cognitive, language, and motor development of the Bayley-III in a sample of children of the first year of life. The psychometric study involved the translation of the original manual and its evaluation forms, analysis of application of evaluation scales in other populations as well as of the structure of the original standardization.

An indirect comparison of the mean scale scores of Russian children with the original American data (utilized for initial standardization), showed no significant differences in the Bayley-III subscales of expressive and receptive language, gross and fine motor subscales in the age range from 2 months 16 days to 10 months 30 days. This result, on the one hand, confirms the quality of translation and adaptation of the original Bayley-III scales, and on the other hand, it speaks in favor of the compliance of the normative boundaries of NPD for Russian and American children. Russian children showed higher scores (0.7 points) on the scale of cognitive development, which could be a consequence of the local characteristics of the experimental sample, in particular, the shift in socio-demographic characteristics towards more educated urban families, and also differences in the characteristics of children with neurological pathology. However, applied difference (effect size) was very low. Taking into account identified patterns, an increase in the size and representativeness of the sample, as, for example, in a study by Chinese authors (Hua et al., 2019), with high probability will lead to full statistical correspondence with the original American data on the cognitive scale.

The findings from the Russian sample of children can be only compared to several studies presented above that carried out the adaptation/psychometric study procedure in other ethnic samples. For example, Ranjitkar et al. (2018) when testing Bayley-III in Nepal, as well as Chinta et al. (2014) in Australia, obtained results that are comparable with the results from the Russian sample only on the fine and gross motor subscales. The study carried out in Denmark (Krogh et al., 2012) showed that the results differ from those obtained in the Russian sample on the subscales of receptive and expressive language, gross motor subscale (lower scores), and fine motor subscale (higher scores) for the children up to 11 months of age; the results on the cognitive scale are comparable.

On the other hand, in contrast to Russian data, the findings from a sample of Dutch children differed significantly across all the scales (Steenis et al., 2015). However, we should note that this is the only study which sample comprised 10 % of children with clinical diagnoses and risks of developmental delay.

It should be pointed out that each of the studies used different comparison procedures with the original Bayley-III data, which could also affect the interpretation of the results. In addition, the following differed: the procedure of adaptation/psychometric study of the measurement tool (performing forward and backward translation, cultural adaptation, adjustments of stimulus materials, etc.), the age of the assessed children, the socio-demographic parameters of the participants, etc. Thus, it is rather difficult to make any conclusions about the comparability of the results obtained in the Russian sample with the data of other populations studies.

In general, the indices of normal NPD obtained in the Russian sample showed almost complete agreement with the original data, which indicates the possibility of the scales application and indirectly confirms the validity of the method even when transferred to another linguistic

and socio-cultural environment. This preliminary result enables us to make positive forecasts for future use of the Bayley-III scales not only as an experimental instrument, but also as a clinical diagnostic tool for assessing NPD. Nevertheless, taking into account the limitations of this sample, the results obtained in our study can be reliably extrapolated only to the population of children aged 2 to 11 months from socially well-educated urban families.

In addition, since the components of the indirect assessment (social-emotional scale and the scale of adaptive behavior) were excluded from the analysis in the present study, the results can only be applied for assessing NPD. In general, for the final confirmation of the psychometric validity of the measurement tool, it is necessary to conduct a large multicenter study, which sample would correspond to all the main socio-demographic characteristics of the population of the RF and proportionally represent different regions, taking into account the balance of the rural and urban population.

References

Aylward, G. P. (1988). Infant and early childhood assessment. In M. G. Tramontana, S. R. Hooper (Eds.), *Assessment issues in child neuropsychology. Critical issues in neuropsychology* (pp. 225–248). Boston, MA: Springer. doi: 10.1007/978-1-4757-9301-7_9

Azari, N., Soleimani, F., Vameghi, R., Sajedi, F., Shahshahani, S., Karimi, H., ... Gharib, M. (2017). A psychometric study of the Bayley Scales of Infant and Toddler Development in Persian language children. *Iranian Journal of Child Neurology*, *11*(1), 50–56. doi: 10.22037/ijcn.v11i1.12056

Bakushkina, N. I., Kiselev, S. Yu., L'vova, O. A., Suleimanova, E. V., & Tuktareva, I. V. (2018). Use of Bayley-III scales to assess neurocognitive development of children in norm and in pathology. *Teoreticheskaya i eksperimental'naya psikhologiya (Theoretical and Experimental Psychology)*, *11*(1), 85–94. (in Russ.).

Ballot, D. E., Ramdin, T., Rakotsoane, D., Agaba, F., Davies, V. A., Chirwa, T., & Cooper, P. A. (2017). Use of the Bayley Scales of Infant and Toddler Development, third edition, to assess developmental outcome in infants and young children in an urban setting in South Africa. *International Scholarly Research Notices*, *2017*, 1–5.

Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development* (3rd ed.). San Antonio, TX: Harcourt Assessment.

Belousova, M. V., & Shvets, E. V. (2019). The influence of information devices and social environment factors on speech development in early age children. *Vestnik sovremennoi klinicheskoi meditsiny (Bulletin of Modern Clinical Medicine)*, *12*(3), 15–20. doi: 10.20969/VSKM.2019.12(3).15-20 (in Russ.).

Byers-Heinlein, K., & Lew-Williams, C. (2013). Bilingualism in the early years: What the science says. *LEARNing Landscapes*, *7*(1), 95–112. doi: 10.36510/learnland.v7i1.632

Chinta, S., Walker, K., Halliday, R., Loughran-Fowlds, A., & Badawi, N. (2014). A comparison of the performance of healthy Australian 3-year-olds with the standardised norms of the Bayley Scales of Infant and Toddler Development (version-III). *Archives of Disease in Childhood*, *99*(7), 621–624. doi: 10.1136/archdischild-2013-304834

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. *Advances in Child Development and Behavior*, *34*, 283–322. doi: 10.1016/S0065-2407(06)80010-8

Colombo, J., & Mitchel, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory*, *92*(2), 225–234. doi: 10.1016/j.nlm.2008.06.002

Fuiko, R., Oberleitner-Leeb, C., Klebermass-Schrehof, K., Berger, A., Brandstetter, S., & Giordano, V. (2019). The impact of norms on the outcome of children born very-preterm when using the Bayley-III: Differences between US and German norms. *Neonatology*, *116*(1), 29–36. doi: 10.1159/000497138

Greenspan, S. I., DeGangi, G., & Wieder, S. (2001). *The Functional Emotional Assessment Scale (FEAS): For infancy & early childhood*. Interdisciplinary Council on Development & Learning Disorders.

Hanlon, C., Medhin, G., Worku, B., Tomlinson, M., Alem, A., Dewey, M., & Prince, M. (2016). Adapting the Bayley Scales of Infant and Toddler Development in Ethiopia: Evaluation of reliability and validity. *Child: Care, Health and Development*, *42*(5), 699–708. doi: 10.1111/cch.12371

Hegde, S., Rao, S. L., Raguram, A., & Gangadhar, B. N. (2013). Cognitive remediation of neurocognitive deficits in schizophrenia. *Neuropsychological Rehabilitation*, 123–153. doi: 10.1016/b978-0-12-416046-0.00007-9

Hoskens, J., Klingels, K., & Smits-Engelsman, B. (2018). Validity and cross-cultural differences of the Bayley Scales of Infant and Toddler Development, Third Edition in typically developing infants. *Early Human Development*, *125*, 17–25. doi: 10.1016/j.earlhumdev.2018.07.002

Hua, J., Li, Y., Ye, K., Ma, Y., Lin, S., Gu, G., & Du, W. (2019). The reliability and validity of Bayley-III cognitive scale in China's male and female children. *Early Human Development*, *129*, 71–78. doi: 10.1016/j.earlhumdev.2019.01.017

Kiselev, S. Yu., Lvova, O. A., & Bakushkina, N. I. (2016). A comprehensive approach to the diagnosis of neurocognitive functions in children at risk of developing autism spectrum disorders. *Izvestiya Ural'skogo federal'nogo universiteta. Seriya 1: Problemy obrazovaniya, nauki i kul'tury (Ural Federal University Journal. Series 1: Issues in Education, Science and Culture)*, *147*(1), 113–121. (in Russ.).

Kiselev, S. Yu., Lvova, O. A., Gliga, T., Bakushkina, N. I., Suleimanova, E. V., Grishina, K. I., ... Martirosyan, S. V. (2016). The assessment of neurocognitive functions in premature infants in the first year of life using Bayley Scales. *Zhurnal nevrologii i psikhiatrii im. C. C. Korsakova (Korsakov Journal of Neurology and Psychiatry)*, *116*(4–2), 62–67. doi: 10.17116/jnevro20161163262-67 (in Russ.).

Kosyakova, O. V., & Bespalova, O. N. (2019). Prevention and therapy of threatened preterm birth in multiple pregnancies. *Zhurnal akusherstva i zhenskikh boleznei (Journal of Obstetrics and Women's Diseases)*, *68*(4), 55–70. doi: 10.17816/JOWD68455-70 (in Russ.).

Krogh, M. T., Væver, M. S., Harder, S., & Køppe, S. (2012). Cultural differences in infant development during the first year: A study of Danish infants assessed by the Bayley-III and compared to the American norms. *European Journal of Developmental Psychology*, *9*(6), 730–736. doi: 10.1080/17405629.2012.688101

Kustova, T. V., Taranushenko, T. E., & Dem'yanova, I. M. (2018). Evaluation of psychomotor development of a child of early age: what should the pediatrician know. *Meditsinskii sovet (Medical Council)*, 11, 104–109. doi: 10.21518/2079-701x-2018-11-104-109 (in Russ.).

Ranjitkar, S., Kvestad, I., Strand, T. A., Ulak, M., Shrestha, M., Chandyo, R. K., ... Hysing, M. (2018). Acceptability and reliability of the Bayley Scales of Infant and Toddler Development-III among children in Bhaktapur, Nepal. *Frontiers in Psychology*, 9, 1265. doi: 10.3389/fpsyg.2018.01265

Rao, U., Hammen, C. L., & Poland, R. E. (2009). Ethnic differences in electroencephalographic sleep patterns in adolescents. *Asian Journal of Psychiatry*, *2*(1), 17–24. doi: 10.1016/j.ajp.2008.12.003

Savostyanov, A. N., Bazovkina, D. V., Saprygin, A. N., Tamozhnikov, S. S., Ausheeva, T. A., Bocharov, A. V., ... Knyazev, G. G. (2018). The psychological and EEG effects of 5-HTTLPR gene polymorphism among people from different ethnic groups in Siberia. In *Cognitive Sciences, Genomics and Bioinformatics (CSGB-2018): Symposium* (p. 32). Novosibirsk: ICG SB RAS. doi: 10.18699/csgb-2018-26

Shifman, B. (2016). The effect of iodine supplementation in pregnancy on early childhood neurodevelopment and clinical outcomes: results of an aborted randomised placebo-controlled trial. *Aktual'naya endokrinologiya (Actual Endocrinology)*. Retrieved from http://actendocrinology.ru/archives/3878 (in Russ.).

Steenis, L. J. P, Verhoeven, M., Hessen, D. J., & van Baar, A. L. (2015). Performance of Dutch children on the Bayley III: A comparison study of US and Dutch norms. *PLoS ONE*, *10*(8), e0132871. doi: 10.1371/journal.pone.0132871

Sun, L., Sabanathan, S., Thanh, P. N., Kim, A., Doa, T. T. M., Thwaites, C. L., ... Wills, B. (2019). Bayley III in Vietnamese children: Lessons for cross-cultural comparisons. *Wellcome Open Research*, *4*, 98. doi: 10.12688/wellcomeopenres.15282.1

Volodin, N. N. (Ed.). (2009). *Neonatology: A national guide*. Moscow: GEOTAR-Media. (in Russ.).

Weiss, L. G., Oakland, T., & Aylward, G. P. (2010). *Bayley-III clinical use and interpretation* (1st ed.). Academic Press.

Yu, Y.-T., Hsieh, W.-S., Hsu, C.-H., Chen, L.-C., Lee, W.-T., Chiu, N.-C., ... Jeng, S.-F. (2013). A psychometric study of the Bayley Scales of Infant and Toddler Development – 3rd Edition for term and preterm Taiwanese infants. *Research in Developmental Disabilities*, *34*(11), 3875–3883. doi: 10.1016/j.ridd.2013.07.006

Yue, A., Jiang, Q., Wang, B., Abbey, C., Medina, A., Shi, Y., & Rozelle, S. (2019). Concurrent validity of the Ages and Stages Questionnaire and the Bayley Scales of Infant Development III in China. *PLoS ONE*, *14*(9), e0221675. doi: 10.1371/journal.pone.0221675

Zavadenko, A. N., Medvedev, M. I., & Degtyareva, M. G. (2018). Assessment of neurodevelopment in children of different gestational age with neonatal seizures. *Zhurnal nevrologii i psikhiatrii im. C. C. Korsakova (Korsakov Journal of Neurology and Psychiatry)*, *118*(11), 35–42. doi: 10.17116/jnevro201811811135 (in Russ.).

No conflict of interest