Моделирование и анализ данных
2020. Том 10. № 1. С. 35–63
DOI: 10.17759/mda.2020100102
ISSN: 2219-3758 (печатный)
ISSN: 2311-9454 (online)

Modelling and Data Analysis
2020. Vol 10, no. 1, pp. 35–63
DOI: 10.17759/mda.2020100102
ISSN: 2219-3758 (print)
ISSN: 2311-9454 (online)

◇◇◇◇◇◇◇◇◇◇◇◇◇**АНАЛИЗ ДАННЫХ**◇◇◇◇◇◇◇◇◇◇◇◇◇

УДК 159.938.3, 159.9.07

Similar or Different? An Item Response Theory Analysis of the Synonyms Test in Adults with and without a History of Institutionalization

Logvinenko T.I.*
Saint Petersburg State University, Saint Petersburg, Russia,
ORCID: https://orcid.org/0000-0001-7430-1963,
e-mail: logvinenkota.spb@gmail.com

Talantseva O.I.**
Saint Petersburg State University, Saint Petersburg, Russia,
ORCID: https://orcid.org/0000-0002-7555-1216,
e-mail: talantcevaoksana@gmail.com

Volokhova E.M**
Saint Petersburg State University, Saint Petersburg, Russia,
ORCID: https://orcid.org/0000-0002-1711-0447,
e-mail: human.nerpa@gmail.com

Khalaf S.****
University of Houston, Houston, United States,
ORCID: https://orcid.org/0000-0002-8809-590X,
e-mail: shiva.khalaf@times.uh.edu

Grigorenko E.L.*****
Saint Petersburg State University, Saint Petersburg, Russia;
University of Houston, Houston, United States,
ORCID: https://orcid.org/0000-0001-9646-4181,
e-mail: elena.grigorenko@times.uh.edu

For citation:
Logvinenko T.I., Talantseva O.I., Volokhova E.M., Khalaf S., Grigorenko E.L. Similar or different? An Item Response Theory Analysis of the Synonyms Test in Adults with and without a History of Institutionalization. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2020. Vol. 10, no. 1, pp. 35–63. DOI: 10.17759/mda.2020100102

***Logvinenko Tatiana Igorevna,** Researcher, Laboratory of the Translational Sciences of Human Development, Saint Petersburg State University, Saint Petersburg, Russia, ORCID: https://orcid.org/0000-0001-7430-1963, e-mail: logvinenkota.spb@gmail.com

****Talantseva Oksana Igorevna,** Researcher, Laboratory of the Translational Sciences of Human Development, Saint Petersburg State University, Saint Petersburg, Russia, ORCID: https://orcid.org/0000-0002-7555-1216, e-mail: talantcevaoksana@gmail.com

The lack of valid and standardized instruments, directed on an assessment of the language domain in adolescents and adults in Russia postulates the urgent necessity of their development. To fill this gap, the language battery, ARFA-RUS, was created and applied in a large project investigating the long-term consequences of raring in institutional care settings on human development. In the current study, an Item Response Theory (IRT) approach was used to examine the psychometric properties of the Synonyms Subtest of ARFA-RUS as the first step of validation of the battery. IRT results demonstrated the test is reliable for the low-to-moderate levels of the assessed ability; yet, to capture a wider ability range, more difficult items are needed. The ARFA-RUS Synonyms Subtest was less suitable for the post-institutionalized group of adults; in this group, the latent ability estimate explained a lower percentage of variance in comparison to adults raised in biological families. With regard to item-specific analyses, two items demonstrated paradoxical patterns with decreased probability of correct response at increased ability. In addition, one item was eliminated from the final version of the Synonyms Subtest due to its poor item fit and low discrimination value.

Keywords: Item-response theory, psychometrics, differential item functioning, language, synonyms, assessment.

***Funding*:** The reported study was supported by the Government of the Russian Federation (grant № 14.Z50.31.0027 “Early deprivation influences biological and behavioral indicators of development”; E.L.G., Principal Investigator). We are grateful to Mei Tan for her editorial support.

******Volokhova Ekaterina Mikhailovna,*** Researcher, Laboratory of the Translational Sciences of Human Development, Saint Petersburg State University, Saint Petersburg, Russia, ORCID: https://orcid.org/0000-0002-1711-0447, e-mail: human.nerpa@gmail.com

*******Khalaf Shiva, PhD,*** post-doctoral fellow at the Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, United States, ORCID: https://orcid.org/0000-0002-8809-590X, e-mail: shiva.khalaf@times.uh.edu

********Grigorenko Elena Leonidovna,*** PhD, professor, principal investigator, Laboratory of the Translational Sciences of Human Development, Saint Petersburg State University, Saint Petersburg, Russia; University of Houston, Houston, United States, ORCID: https://orcid.org/0000-0001-9646-4181, e-mail: elena.grigorenko@times.uh.edu

1. INTRODUCTION

1.1 Language Skills in Institutionalized Individuals

Language is a fundamental human ability, and it is a fundamental component of many different skills and processes (i.e., memory, executive functions, learning). Like other complex skills, language development is influenced by a variety of genetic and environmental factors and their interactions. One of the key risk and protective factors of language development is linguistic input, defined as an obligatory (i.e. experience-expectant) element of the environment for every individual, as well as access to a caregiver, proper nutrition, and cognitive and sensory stimulation [23, 70]. However, not everyone receives a sufficient amount of linguistic input. Namely, an individual’s linguistic input can be placed on a continuum that ranges from a lack of exposure in an environment to a rich and stimulating environmental experience.

The consequences of the absence of linguistic input were investigated in studies of extreme language deprivation, such as children who were sensorily deprived due to inborn deafness [42] or full social isolation [19]. More subtle, but, correspondingly, more widespread, are cases of children who have not had caregivers to provide them with the necessary linguistic input. Orphaned, abandoned, maltreated and underprivileged children are usually deprived of a rich linguistic environment while being reared in state institutions or being raised in families with low-quality caregiving [70].

Longitudinal and cross-sectional studies have repeatedly examined the relationship between institutionalization and language deficits [7, 20, 21, 58]. Particularly, the following language domains have been reported to be at risk: expressive language skills by the age of 42 months [71]; expressive and receptive vocabulary, and narrative skills by 4-5 years of age [3]; sentence repetition, nonword repetition, and word identification by the age of 8 years [72]; performance on the Comprehensive Assessment of Spoken Language and Clinical Evaluation of Language Fundamentals (Fourth Edition) by age of 8-11 years [38]; sentence comprehension by 8 years of age [14]; and vocabulary skills by the age of 11 years [45]. Studies of institutionalization and their results are very heterogeneous due to variability in the age of placement and the time spent in institutions, quality of institutional settings, following remediation or adoption, transition to different language environments, and so forth. However, despite sample and assessment differences, language abilities of institutionalized individuals are expected to be lower than the population mean in a wide range of domains, including the core language functions described above.

1.2 Vocabulary Skills

Vocabulary (including vocabulary skills, vocabulary knowledge, and lexicon), broadly defined, is the understanding of words and their meanings. Vocabulary indexes the broader ability – lexical skills and processing, which refer to what one knows about a word, its usage, its components – and the word's phonological, orthographic, and physical forms – and their link to its mental representation [48]. One of the most commonly used classifications of vocabulary skills belongs to Kate Nation [22], who distinguishes receptive vocabulary as comprehending a word in listening and reading, and productive vocabulary as producing a word in speaking and writing.

Vocabulary skills are essential for all core language abilities. Studies have repeatedly shown the association between size and quality of early vocabulary and later development of listening, writing, reading, and speaking skills [5, 27]. This relationship is persistent in adulthood. Andringa and colleagues [1] showed the intuitively comprehensive connection of better vocabulary knowledge with increasing listening comprehension. Moreover, it has been shown that participants' lexicon size impacts accurate language perception, specifically in the recognition of orally presented words [32], words in lexical decision tasks [73], and speech recognition in adverse conditions [6]. The involvement of vocabulary skills in language production is well investigated, namely in speeded pronunciation, picture naming, and verbal fluency tasks [57, 59, 64, 73]. Similarly, knowledge of word meanings mediated by listening comprehension affects text comprehension in children, adolescence, and adults [51] and vice versa, since a sufficient part of vocabulary learning is done through reading [13, 28].

Being a crucial component of general language abilities, vocabulary skills are a strong indicator of an individual's achievements, especially in education (e.g., [46, 49, 60]). Correspondingly, vocabulary deficits put academic, vocational, and mental health outcomes at risk at various stages of life [2]. Unfortunately, the discrepancies in vocabulary abilities between individuals emerge primarily in childhood and are highly dependent on family and socio-economic background [8, 18]. The research of Hart and Risley [25] demonstrated that by the age of three the amount of words heard by children in high-income families is double that of disadvantaged children. These differences in the quantity of input influence the richness of children's vocabulary, and hence, their later school success [26]. We expect similar effects on individuals raised in institutional care in Russia. First, because the lack of linguistic input has been documented in Russian institutions [47, 63]; and second, because the lexical deficiencies of institutionalized children have been identified worldwide (e.g. [3, 38, 45]). However, the later-life trajectories of language development in institutionalized individuals are largely unknown in Russia. A simple explanation could be that there is a lack of diagnostic assessments of vocabulary and other language-related skills for adults. Such an instrument is especially important for proper diagnosis and the remediation of individuals' language skills.

1.3 Measures Used to Assess Vocabulary

There are several measurement approaches, that examine vocabulary skills. According to McGregor and colleagues [44], there are two directions, namely "breadth and depth", by which one can arrange lexical skills. Vocabulary breadth mainly refers to how many words a person can produce or identify. On the other hand, vocabulary depth requires using words in the whole language structure.

There are multiple standardized tests for adolescents and adults that assess both vocabulary breadth and depth skills in English. To assess vocabulary breadth, one can use word-picture matching tasks (for example, the Peabody picture vocabulary test-fifth edition [16]) or picture naming approaches (like the Expressive vocabulary test-third edition [69]). On the other hand, to investigate depth of word comprehension, word definition tests can be used (i.e., vocabulary Subtests of the Wechsler intelligence scale for children (WISC)-fifth edition [66] or the Wechsler adult intelligence scale (WAIS)-fourth edition [65]). However, in Russia, there is a lack of standardized psychometric tests that evaluate lexical skills in adolescent and adult populations. Russian versions of the WAIS, including its verbal Subtests, were not standardized based on Russian norms and are outdated. The Russian Passive Vocabulary and Active Vocabulary Subtests [75] are designed for 3-9-year-old children and thus are not applicable to adults. While the Russian Aphasia Test (RAT) [31] that is currently in the process of standardization and normalization includes Subtests that assess word production and comprehension skills, it was made for a population with language impairments and fails, and therefore may not address the general population in the best way.

In the middle of the vocabulary test continuum, located along the "depth" and "breadth" axis, there are synonym and verbal analogy tests. In such tests, one does not need to define the word but has to understand the meaning of the word and other linguistic characteristics (for example, morphological). In verbal analogy tests there are usually three stimulus words: two of them illustrate a logical relationship rule, and for the third, the participant finds a word to complete the analogy, based on the rule between the words in the example. In tests of synonyms, the participant needs to produce or to choose a word with the same meaning as the stimulus word.

There are multiple psychometric batteries, tests of school performance, and aptitude tests that include analogy and synonyms subtests in English. Some of the common current English verbal analogy and synonym tests for adolescents and adults are listed in Table 1. These tests are commonly used to asses both children and adults because the execution time is short, and they simultaneously assess two verbal cognitive abilities, namely, vocabulary skills and reasoning.

Disappointingly, to the best of our best knowledge, Russian clinical language assessments do not include verbal analogy or synonym tests. Existing tests that include such tasks assess academic progress (e.g., Edinyy Gosudarstvennyy Ekzamen [Unified State Exam]) and are not designed for the assessment of individual differences. The lack of Russian standardized diagnostic instruments for assessing the language domain in adolescents and adults make it extremely important to develop a valid and reliable language and speech assessment tools, capable of differentiating problems in different language subdomains.

1.4 ARFA-RUS Synonyms Subtest

To fill the gap in Russian psychodiagnostic tools of language assessment in adults and adolescents, the ARFA-RUS Synonyms Subtest was developed. This subtest is part of a large ARFA-RUS battery developed in the Laboratory of the Translational Sciences of Human Development of Saint-Petersburg State University. ARFA-RUS was applied in a large research project studying the bio-behavioral features of the development (with a focus on the language domain) of adult orphanages raised in institutional-care settings. Some data on ARFA-RUS Synonyms Subtest performance as a part of this large research project have already been reported[35][1].

The ARFA-RUS Synonyms Subtest resembles other synonyms tests with respect to its structure. The aim of this task is to recognize and distinguish all synonymous words from words that are not synonyms. Also, the ARFA-RUS Synonyms Subtest partially resembles verbal analogy tests because it contains an example of synonymous words that provide a model of a correct response. The items on the ARFA-RUS Synonyms Subtest are formulated as yes-no questions, which differentiates this Subtest from listed (Table 1) English synonyms and verbal analogy tests that contain either open or multiple-choice format tasks.

1.5 Item Response Theory (IRT) approach

In modern test theory, item response theory (IRT) is the dominant psychometric paradigm for scale development, analysis, and scoring. IRT assesses the extent to which items evaluate individual differences on some specified construct. Estimates of IRT model parameters provide information about the discriminating power of an item separately from its

[1] In the publication by Kornilov et al. [35], the original name of the ARFA-RUS Synonyms Subtest, that is, "Analogies", was used. We decided to change the name of the subtest to Synonyms because the new name better reflects the task and structure of the test.

difficulty or severity [61]. Traditionally, IRT models are named according to the number of parameters that are used to model the characteristics of an item. If a single characteristic of the item is modeled (e.g., item difficulty), the IRT model is called a 1-parameter model (1PL). If two characteristics of the item are modeled (e.g., item difficulty and item discrimination), the IRT model is called a 2-parameter model (2PL). When three characteristics of the item are modeled (e.g., item difficulty, item discrimination, and item pseudo-guessing, which is common for a multiple-choice item test), the IRT model is called a 3-parameter model (3PL). In general, IRT provides information about measurement precision across the range of a latent trait at both the item and test level rather than providing only a single reliability estimate for all participants. In the case of the current study, the latent trait under investigation is the ability to distinguish similar from discrepant word pairs. Table 2 summarizes the terminology used in this article in connection with IRT.

The advantages of IRT can be clarified by comparing this approach with Classical Test Theory (CTT). By CTT we are referring to traditional psychometric methods (such as factor analysis and Cronbach's α) based on the idea that a person's observed or obtained score on a test is the sum of his/her true score (error-free score) and error score. CTT concentrates on the entire test, rather than on the responses to items. While CTT methods are easy to compute and interpret, they have some principal limitations, including the (1) assumption of a linear association between the measured latent trait and obtained scores, which rarely corresponds to empirical reality in the case of psychological constructs; and (2) dependency of reliability parameters from sample characteristics and sample size. In contrast to CTT, IRT requires significantly more time and computational resources, large sample sizes for analysis, and the need for stronger assumptions. However, the following benefits of using IRT outweigh its complexity: (1) assumes a nonlinear relationship between the latent trait and test scores, accounting for the random nature of these responses and use of probabilistic models to explain their distribution; (2) provides trait scores at the item level; (3) provides the reliability of each item at different levels of the latent trait, controlling for the characteristics of the items in the scale (e.g., difficulty, discrimination, etc.), that can be especially useful in the identification of items that may contribute little or make distortions to measurement precision; (4) allows for the presence of separate parameters for the effects of the subjects' abilities, skills, or attitudes and the properties of the item; (5) permits independent examination of psychometric properties from sample characteristics [17, 37, 50, 54].

Within the IRT approach, a common practice is to investigate the potential bias of the test and/or items for various subgroups using the differential item and test functions (DIF and DTF, respectively). DIF/DTF occurs when an item/test measures a latent construct differently for one subgroup of a population than it does for another [36, 40]. DIF might be a threat to test validity for one or several subgroups. However, it is not necessarily that the item or test is unfair, because DIF indicates the presence of a latent trait, which might or might not be meaningfully or intentionally related to the targeted construct [41]. In the current study, we will examine the potential bias of test items for two groups of individuals, namely, adults with and without a history of institutionalization. The investigation of the test bias is essential because the Synonyms Subtest of ARFA-RUS will largely be used for incarcerated adults in future studies, thus the reliability of the items for this subgroup is of the utmost importance.

1.6 Aims

The first aim of the study is to evaluate the psychometric properties of the Synonyms Subtest of the new language battery ARFA-RUS using item response theory (IRT). The second aim is to investigate the potential test and items bias for two groups: individuals who experienced institutional care (IC) and individuals raised in biological family care (BFC). We expect the test to work equally fairly for both groups when we control for the overall differences in the measured ability. Although our main objective is to examine the psychometric properties of the ARFA-RUS Synonyms Subtest, we also demonstrate practical application of IRT models for other areas of cognitive and personality assessments. For this, the decision-making process and analysis are transparently described.

2. METHOD

2.1 Participants

The initial sample for the 1PL and 2PL models included 655 native Russian-speaking individuals who took part in a larger study of the long-term effects of institutionalization (supported by the Government of the Russian Federation). For the group analysis, 5 participants were excluded due to missing information about group, sex, age or other demographic characteristics. The final sample for group analysis consisted of 650 participants (366 females, 284 males) ranging from 15 to 38 years old (*Mdn* = 19, *M* = 20.38, *SD* = 4.703). Of these, 342 had experienced institutional care (IC) and 308 were raised in biological family care (BFC). The inclusion criteria for the IC group were: age ranging from 16 to 35 years old, being native Russian language speakers, and having experienced living in an orphanage or related institution. Exclusion criteria were the presence of: cerebral palsy, epilepsy, brain diseases, genetic abnormalities, systemic autoimmune diseases (e.g. hypothyroidism, systemic lupus erythematosus); metabolic disorders (e.g. metabolic syndrome, hypertension, fibromyalgia), serious brain injuries (e.g., brain surgery; trauma with loss of consciousness) or any traumatic brain injuries in the last 6 months. The BFC group was matched by sex, age, socioeconomic status and educational level. Participants for the IC group were recruited through a number of vocational schools and organizations that support adolescents and young adults left without biological parental care and raised in institutional settings (baby homes and orphanages), located in two large cities in the Russian Federation. Participants for the BFC group were recruited via colleges and social networks.

2.2 Assessment

The ARFA-RUS Synonym Subtest was presented as part of the ARFA-RUS battery. The Synonyms Subtest assesses vocabulary knowledge and the ability to understand synonymy relationships between associated words using a yes-no format. The Subtest includes 25 items. The participant is asked if two presented words have the same meaning or not, and are asked to label them as “Similar” or “Different”. An example of a pair of words that have similar meanings is also provided (смелый-бесстрашный [smely`j-besstrashny`j], eng. brave-fearless). The two words in each item always represent the same part of speech (noun or adjective) and may or may not have any relationship to each other. Relationships may belong to such semantic classes as synonyms (антропогенный-человеческий [antropogenny`j-chelovecheskij], eng. anthropogenic-human), antonyms (комизм-трагизм [komizm-tragizm], eng. comic-tragedy), or affiliation to the same category (меридиан-экватор [meridian-e`kvator], eng. meridian-equator).

Participants were assessed individually; the Synonym Subtest was always the second subtest of the administered battery. It was completed on paper forms by participants, with no time limit. Instruction was provided both in writing (on the paper forms) and auditorially.

2.3 Analytical procedure

All analyses were conducted in the *R* programming environment [52]. The analytical plan was as follows: 1) describe the test and item characteristics across and within the two study groups; 2) investigate the dimensionality of the test; 3) compare the 1PL and 2PL models regarding their fit to the data; 4) perform IRT group analysis for the BFC and IC groups; and 5) examine the differential item functioning for these groups.

The latent dimensionality of the test was examined with exploratory factor analysis. The appropriateness of the factor analysis for the data matrix was verified by the Kaiser–Meyer–Olkin measure, with a sampling adequacy of .86.

To perform the IRT analyses, we used the *ltm* [56] and *mirt* [11] packages. The fit of the models was based on the marginal maximum likelihood / EM (expectation-maximization) algorithm (e.g., [9]). Marginal maximum likelihood is one of the basic estimation techniques in IRT, along with maximum likelihood, conditional maximum likelihood, and the Markov chain Monte Carlo method. The key characteristic of marginal maximum likelihood is the assumption that the participants are randomly sampled from a larger distribution. The estimation is conducted by maximizing the observed data loglikelihood, which depends on latent data and integrating the person effect out of the joint likelihood [33]. It utilizes the approximation of required integrals; to accomplish this approximation, the Gauss-Hermite quadrature rule was used.

Item analysis began with fitting a 1PL model. For each item, the difficulty parameter (b) was estimated and the slope (discrimination) parameter (a) was constrained to be equal. For the 2PL model, the slope (discrimination) parameters and the difficulty parameters were estimated for each item. The terms “easy”, “moderate” and “hard” were used to indicated relative item difficulty. For the interpretation of the slope, as per Baker [4], the following labeling was used: none – 0; very low – 0.01–0.34; low – 0.35–0.64; moderate – 0.65– 1.34; high – 1.35–1.69; very high – >1.70; perfect – +infinity.

To choose the best fitting model, the fit indices were investigated: lower values of Akaike information criterion (AIC), Bayesian information criterion (BIC), log-likelihood and higher values of test information were preferable.

A Monte Carlo procedure was used to approximate the distribution of the item-fit statistic. The statistical significance of the test indicates that the null hypothesis should be rejected indicating that the items do not fit the model.

To detect potential test bias, we applied several DIF-methods (*difR* package [40]), which differed on how the participants were matched on the measured ability. The analyses included classical methods – the Mantel-Haenszel statistic, standardization, and logistic regression – which matched students based on the total scores; IRT-based methods, such as the Wald $\chi 2$ test (also known as Lord's test [39]) and Raju's area test [53], which consider student ability as a latent variable, estimated together with item parameters in the model. Those five methods were applied to compare the results, however, we decided to use the Wald $\chi 2$ test of IRT-based methods group as the reference. The first reason is that IRT-based methods are more accurate than classical ones since they measure the latent trait instead of the total score. The second reason is that they allow detecting non-uniform DIF [41].

3. RESULTS

3.1 Descriptive statistics

Table 4 shows the mean, standard deviation, Cronbach's alpha, skewness, and kurtosis values for the global score of the ARFA-RUS Synonyms Subtest, separately for the IC and the BFC groups. An independent samples t-test found that participants from the BFC group scored significantly higher than the IC group (t (568) = -19.11, p <.0001) with a large effect size (d = -1.61). Cronbach's alpha for the global score was acceptable (.80), and was also found to be acceptable for the BFC group (.74), but was poor for the IC group (.58). This issue will be thoroughly examined in the IRT framework since CTT does not clarify between-group differences. Skewness absolute values were <|1|, and kurtosis absolute values were <|3|, demonstrating sufficient univariate normality. Item descriptives may be found in Supplemental Table 1. As the scale reliability could not be notably improved through any item exclusion, it was concluded that all items were performing adequately. The frequency of correct responses on items varied from 41% to 87%. Mean values for items 1, 3, 9, and 23 exceeded 0.8; thus, these items were determined to be the easiest and should be examined more closely in the IRT analysis. For items 2, 13, 18, and 24, correct responses were less likely, but approximately equal, compared to the likelihood of wrong answers; these items were determined to be the most difficult.

3.2 Dimensionality analysis

First, we conducted Horn's [29] and a modified parallel analysis [15], which yielded equivocal results, with the former indicating unidimensionality and the latter pointing to the possibility of a second dimension. Therefore, we investigated one- and two-factor exploratory factor solutions. In the one-factor model, the factor eigenvalue was 3.79, which explained 15.2% of the variance in the data. In the next model, two factors cumulatively explained 18.2 % of the variance, with 11.1% uniquely explained by the first factor. Moreover, the examination of factor loadings and the corresponding items revealed that items were clustered together based not on test but method factor. Specifically, the first factor was indicated by items for which the response "similar" was correct, while the second factor included items requiring the "different" response. Since we cannot claim that the cognitive processes are different when judging those two item types, we considered the second dimension as redundant and further analysis was conducted using unidimensional solutions.

3.3 IRT-model evaluations

Table 5 shows the model fit statistics for the 1-PL, 2-PL, and corrected 2-PL models, and the difficulty and discrimination threshold parameters for each item of the ARFA-RUS Synonyms Subtest.

The 2-PL model was superior to the 1-PL model on the basis of the BIC, AIC, log-likelihood and Likelihood-Ratio (LR) test ($\chi^2(24) = 309.82$, $p < .001$). In both the 1-PL and the 2-PL models, some items had a poor fit, but the number of misfit items in the 2-PL model was three times less than the 1-PL (see item fit indexes in Table 5). The correlation between the difficulty parameter of the 1-PL and 2-PL models was high ($r = 0.93$). Thus, although the 1-and 2-PL models seemed not to differ much with respect to difficulty, the 2-PL model could be considered the more appropriate model to describe the ARFA-RUS Synonyms Subtest.

In the 2-PL model, all items but 11 and 18, fit the model. These poorly fitting items need to be examined for modification or removal. The difficulty of the items ranged between easy and medium; item 23 was the easiest, while item 24 was the hardest (b = -2.34 and 1.07, respectively). The discrimination parameter was very low for item 12; low for items 13, 15, 18, 24; high for items 5, 9, 20; very high for 3, 14, 19; and moderate for the remaining items. The higher discrimination values indicate that those items are better at discriminating among individuals with low and high latent ability (i.e. the ability to discern similar from discrepant word pairs).

The ICCs of all of the items (Figure 1) indicate the expected increasing probability of responding correctly at higher levels of ability. For the low discrimination items, this increase was shallower, which means that these items differentiate between persons of high and low ability less effectively, that is, persons with different ability levels are likely to respond similarly to these items.

The information function along with the standard error measure for the overall ARFA-RUS Synonyms Subtest demonstrates that the test is most reliable at low-to-moderate levels of the latent trait (right panel of Figure 1). The most information is provided for individuals with ability levels of about –1.0. At this point, the standard error is the lowest and, conversely, the information is highest, indicating that the test should be considered easy and less able to discriminate individuals at higher ability levels.

Within the CTT framework the exclusion of item 18 did not result in the improved reliability of the entire test as indicated by the Cronbach's' α statistic. While analyzed within the IRT framework, this item demonstrated poor item fit and low discrimination, therefore it was removed from the final version of the subtest. This decision was also influenced by concerns regarding the features of the stimuli words: dekada [decade] - stoletiye [century]). Both words belong to the same semantic category, and differ not qualitatively but rather reflect quantitative differences of the time periods, in contrast to other stimulus pairs.

Vuong's test for non-nested models demonstrated that excluding of item18 led to a significant improvement of the 2-PL model ($z = -86.53$, $p < .001$), which was also confirmed by advancing BIC, AIC, and log-likelihood indexes (Table 5).

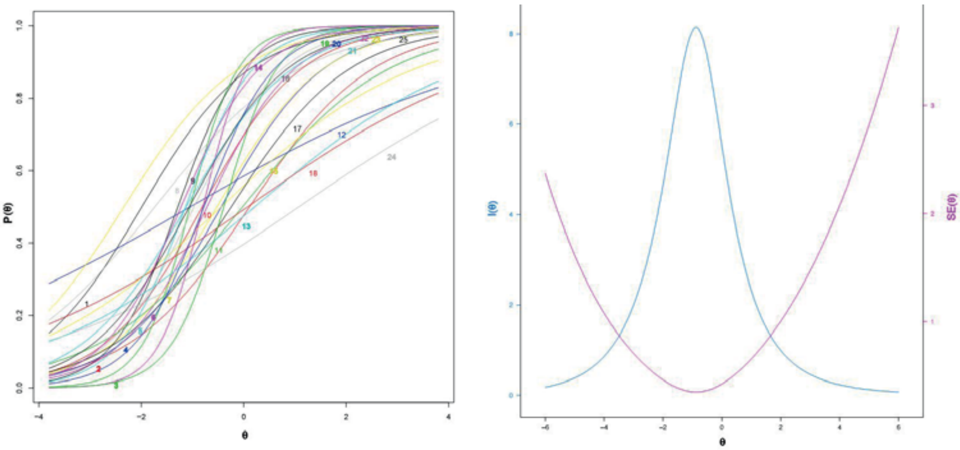

Figure 1. Item characteristic curves for the 2-PL model (left panel) and the test information function with standard error measure for all items of the 2-PL model without item 18 (right panel). The ICCs are highlighted by different colors and numbers. The X-axis represents latent ability levels, the Y-axis (left panel) – probability of the "correct" response, the Y-axis (right panel) – information (blue line) and standard error (pink line)

3.4 Differential test and item functioning

The examination of the differential test and item functioning started with a fully unconstrained model, that is, the discrimination and the difficulty parameters were estimated for each item respective to the group. The analyses revealed little agreement among DIF-detection tests. Still, all of the items were identified as functioning differentially by at least two tests (Supplemental Table 2). However, when none of the items are constrained, the presence of DIF items might influence the estimation of the parameters for non-DIF items, with the latter being wrongly identified as DIF [40]. Namely, without proper item parameters equating among groups, the analysis does not just identify DIF, but might reflect a combination of differential item functioning and latent-trait distribution effects. Therefore, a set of anchor items that will be constrained was needed. For this, we used the unconstrainted model as a reference and compared it via likelihood-ratio tests with the models where group equality constraints were added to the difficulty and discrimination parameters for each item one by one (see [12]). Thus, eleven items were found to be invariant and were used as anchors. There was no statistically significant difference between the final multigroup model with anchors and the fully unconstrained model ($\chi2$ (*df*) = 19.48 (20), p = .491), proving that the anchor usage did not dramatically change the estimation.

The final model demonstrated adequate fit based on the M2* family of statistics with M2*(*df*) = 890.857, CFI = .808, RMSEA = .035, and an SRMSR of .072 and .071 for the BFC and IC groups, respectively. Note that the CFI value is slightly lower than the suggested thresholds [30], which might indicate the preference for a more restricted model. The reference latent mean and variance in the BFC group were fixed to 0 and 1, respectively. The estimated latent mean for the IC group was -1.622 with the variance 0.378, suggesting that the IC had a notably lower estimated ability on the test and more homogeneous responses. Additionally, whereas the latent factor accounted for 26.4 % of the BFC group scores (eigenvalue = 6.34), it explained half of the variance of the IC group scores (12.5 %, eigenvalue = 3.00). The Wald χ^2 test revealed the presence of DIF in six items. The DIF test results, as well as item parameters for each group, are presented in Table 6. Non-uniform or crossing DIF was also detected. That is, some items favored the BFC group on one level of latent ability and the IC group on another. For items 1, 8 and 23, the probability of correct responses was above 0.5 even on the lowest ability levels for the IC group. Furthermore, the trace lines for items 11 and 12 had negative slopes for the IC group; in other words, the probability of the correct answers decreased for higher ability participants in the IC group, indicating non-monotonicity of their responses to these items. The above-mentioned effects are displayed in Figure 2.

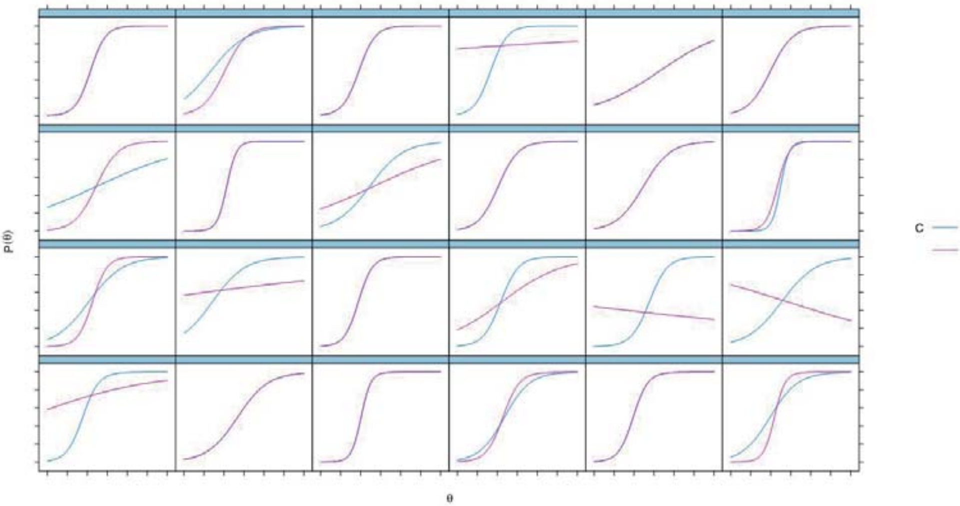

Figure 2. Item characteristic curves for the BFC and IC groups. The X-axis represents latent ability levels, the Y-axis – probability of the "correct" response. Items 2, 3, 5, 9, 14, 16, 17, 20, 22, 24, 25 were used as anchors, therefore their parameters are equivalent across groups. Item 18 was eliminated in the previous analysis steps

The practical implementation of DIF analysis is related to the question of whether any specific item demonstrates a bias favoring a specific group; dropping such an item will improve test fairness for groups, disadvantaged by such an item. Importantly, one can investigate the differential functioning of the whole test (DTF); even if some items demonstrate DIF, it does not always lead to overall test bias (e.g., DTF). The omnibus tests revealed that the average amount of test scoring bias (signed DTF) between the response curves of the two groups was 0.35 of a raw test score (1.43 % of the total score) favoring the BFC group, while the absolute deviations in item properties over the test (unsigned or total DTF, which disregards the bias towards the specific group) was 1.38 points (5.76 % of the total score). The expected total scores for both groups and the illustration of the changes in the overall test bias across all ability levels are in Figure 3. The results indicate that at the lower end of ability, the ARFA-RUS Synonyms Subtest favors the IC group, but from the theta level of -1.4 to the highest ability levels, the test favors the BFC group.

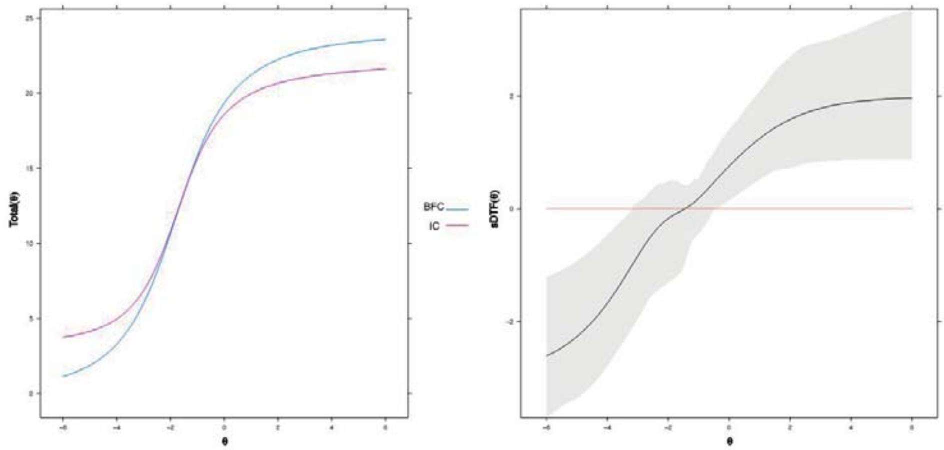

Figure 3. Expected total test score for the IC and BFC groups (left panel), and differential test functioning (right panel). The X-axis represents latent ability levels, the Y-axis (left panel) – score on the ARFA-RUS Synonyms Subtest (max = 24), and the Y-axis (right panel) represents sDTF – signed differential test functioning. Negative values of sDTF indicate that the reference (BFC) group scores lower than the focal group (IC), positive values indicate the opposite

3.5 Discussion

The implementation of standardized and valid instruments is crucial for both practice and research purposes in studies of individual differences and clinical research. The lack of reliable standardized methods for assessing language abilities in Russia has encouraged researchers to develop new approaches from scratch. Once the test is developed, however, its psychometric properties must be evaluated.

The ARFA-RUS Synonyms Subtest was created in the Laboratory of Translational Sciences of Human Development of Saint-Petersburg State University as a part of the ARFA-RUS language battery. The Subtest was used to investigate the lexical skills of participants with a history of being reared in institutional care settings, as well as their peers raised in biological families. In the current study, we addressed the reliability of the ARFA-RUS Synonyms Subtest and investigated to what extent the Subtest can be used in the study groups without DTF.

First, to determine the applicability of the ARFA-RUS Synonyms Subtest for research and practice we evaluated its psychometric properties using item response theory (IRT). The dimensionality check revealed some evidence of a second dimension. In the two-factor exploratory model, one factor was formed by items requiring the answer “similar”, while the second factor – by items requiring the “different” response. There is evidence that the identification of synonyms and antonyms measure one domain of cognitive abilities, yet a small proportion of the variance is explained by a method factor [67]. Thus, in this article the unidimensional model was used, however, future work should specifically prove that the latent construct is independent of the selected method of measure. Subsequently, we estimated 1-PL and 2-PL models, and the latter was preferred based on the model fit indices and likelihood comparison. We eliminated one item because it had a poor fit, low discrimination, and, more importantly, raised concerns regarding its content. The final 2-PL model with 24 items had an adequate fit; all except two items fitted the model well; the discrimination parameter varied with five items demonstrating low discrimination power and others moderate-to-high discrimination. The item difficulty ranged between easy and medium levels (from -2.38 to 1.07). In general, the IRT demonstrated that the test is reliable at low-to-moderate levels of ability to understand the relationship between synonyms (i.e., the ability to understand the differences associated with similar versus discrepant word pairs); this observation corresponds to the notion of acceptable internal consistency of the global test scores calculated with CTT. However, to capture a wider ability range, more difficult items are needed.

The second aim was to investigate the equality of the test and items for the two subgroups, namely, the average adults raised in biological families and adults who experienced living in institutional care. The IRT analyses discovered that adults with a history of institutionalization had a lower latent mean, and the presence of the latent factor explained the considerably lower percentage of the variance in their scores in comparison with the reference group. Moreover, the adults with a history of institutionalization scored more unfavorably on the test, that is, received negatively biased test scores. However, the bias was inconsistent: at the lower end of ability, the Subtest favored the institutionalized group, but from the theta level of -1.4 to the highest ability levels, the test favored the reference group. Specifically, we detected six non-uniform or crossing differentially functioning items. That is, these items were unequal for the two subgroups, even when their differences in the latent ability were controlled. Furthermore, items 11 and 12 revealed paradoxical patterns: the decreased probability of correct response at higher ability levels. Since those items worked well for the reference group of average adults, we are not going to modify them; however, they should be removed when used for special populations or non-average subgroups. Overall, the ARFA-RUS Synonyms Subtest demonstrated DTF and was found to be less appropriate for individuals with early experience of institutionalization (i.e., individuals whose early language input might have been challenged). The trend was also reflected by a poor Cronbach's alpha value in the post-institutionalized group in contrast with acceptable indices for the reference group.

We hypothesize that the group discrepancy may be due to the existence of a second latent trait that was unintentionally measured with our construct of interest. To successfully perform verbal analogies tests, different components of executive functions are needed [55]. Executive functioning is known to be altered in post-institutionalized children [43], and those alterations might be persistent; therefore, we assume that executive functions may be an uncontrolled latent trait in adult samples, too. However, to explore this assumption, further investigations are needed.

Also, since the group of post-institutionalized adults had, on average, lower ability in the synonymy identification than the reference group, they could have used the guessing strategy more frequently to solve the task. The chance answers might partially explain the inconsistent difficulty-to-probability item curves in that group. The limitation of the current study is that we restricted the IRT analysis to the one- and two-parametric models, however, a future direction of the study might be to use a more advanced model with guessing parameter (3-PL) to investigate the issue in greater detail.

The IRT approach to psychometric analyses has been used worldwide for decades and has proven its effectiveness. Although it has provided several advantages over traditional psychometric approaches, it is not a routine part of test development and evaluation in Russia. It is possible that IRT is rarely chosen because it requires more statistical knowledge and large sample sizes, and in psychology and related fields, investigators might intentionally or unintentionally choose confirmatory factor analysis over IRT to appraise the psychometric characteristics of the instruments they develop ([74]; see [62] and [34] for CFA and IRT comparison). In the current study, all steps were described as transparently as possible to encourage the utilization of IRT for further research in the field of test development and psychometric evaluation in Russia.

Table 1

English Verbal Analogy and Synonyms tests suitable for the assessment of adolescents and adults

Group of tests	Test, *Subtest*	Description
Verbal Analogy	Clinical evaluation of language fundamentals – Fifth Edition. CELF-5, *Word Classes* (Wiig et al., 2013) [68]	Age range: 5:0 – 21:11 Task format: multiple-choice Task: to identify two words (or pictures) that are related by semantic features, function, place, or time from a list of four words
	Test of Adolescent and Adult Language – 4th Edition. TOAL-4, *Spoken Analogies* (Hammill et al., 2007) [24]	Age range: 12:0 – 24:11 Task format: open Task: to finish an examiner's partial analogous sentence with a word to complete the analogy.
Synonyms	Test of Adolescent and Adult Language – 4th Edition. TOAL-4, *Word Similarities* (Hammill et al., 2007) [24]	Age range: 12:0 – 24:11 Task format: open Task: to write a synonym (correct spelling is irrelevant) for a printed stimulus word.
	Comprehensive Assessment of Spoken Language – Second Edition.CASL-2, *Synonyms* (Carrow-Woolfolk, 2017) [10]	Age range: 3 – 21 Task format: multiple-choice Task: to identify, using a list of four words, a word synonymous to the target word.
	Expressive Vocabulary Test-Third Edition. EVT-3, *last 52 items of the test* (Williams, 2018) [69]	Age range: 2:6 – 90+ Task format: open Task: to provide a synonym to the presented word or a spoken phrase

Note: Age range – range of ages for which the test is applicable; Task format – format of questions; Task – brief description of the task.

Table 2

Item Response Theory: key terms and definitions

Term	Definition
Latent ability/trait (theta, θ)	IRT model parameter that indicates unobservable construct (latent trait) being measured by a scale.
Item difficulty parameter (b, β)	IRT model parameter that indicates the difficulty (or severity) of an item response. β
Item discrimination or slope parameter (a, α)	An item parameter that indicates the strength of relation between an item and the measured construct or latent trait (θ). The slope parameter is usually positive (negative values mean the positive responses to the item are associated with lower values of the trait being measured, which usually means the item is keyed incorrectly).
Information function/curve	A function that indicates the range over the construct being measured (θ) for which an item or scale is most useful for discriminating among individuals.
Differential item functioning (DIF)	DIF occurs when subgroups from a population perform differently on an item after controlling for the overall differences between subgroups on the latent trait that was measured. There are two types of DIF. The uniform DIF represents the difference in success probabilities of subgroups that is constant across the latent ability levels. That is, an item consistently gives one subgroup an advantage. If the subgroup differences are not constant across ability levels but depend on it, that is the nonuniform or crossing DIF.
Unidimensionality assumption	Assumes that one underlying (or dominant) factor accounts for a person's response to a question within the scale.

Table 3

Sample characteristics

Variable		IC	BFC	Overall
Overall		308	342	650
Age (years)		M = 19.13, SD = 4.08	M = 21.51, SD = 4.94	M=20.38 SD = 4.70
Entrance Age	0-4 years	67 (19.59%)	–	–
	5-12 years	147 (42.98%)	–	–
	3-16 years	81 (23.68%)	–	–
	NA	13 (3.80%)	–	–

Variable		IC	BFC	Overall
Duration of institutionalization	0-2 years	36 (10.53%)	–	–
	>2 years	255 (74.56%)	–	–
	NA	17 (4.97%)	–	–
Education	higher education and uncompleted higher education	6 (1.94%)	70 (20.47%)	76 (11.69%)
	secondary education and uncompleted secondary education	157 (50.97%)	129 (37.72%)	286 (44%)
	secondary professional education	129 (41.88%)	127 (37.13%)	256 (39.38%)
	NA	16 (6.19%)	16 (4.68%)	32 (4.92%)
Educational program during school	standard	233 (75.65%)	338 (98.83%)	571 (87.85%)
	corrected for special educational needs	75 (21.93%)	3 (0.88%)	78 (12%)
	NA	0 (0%)	1 (0.29%)	1 (0.15%)

Note: Entrance Age – age of placement in institutional care settings, M – mean, SD – standard deviation, IC – institutional care group, BFC – biological family group, NA – missing value.

Table 4

Mean, standard deviation, Cronbach's α, skewness, and kurtosis values for the ARFA-RUS Synonyms Subtest

Variable	IC	BFC	Overall
Mean	13.21	19.08	16.44
SD	3.63	3.71	4.71
Range	5-25	9-25	5-25
Cronbach's alpha	.58	.74	.80
Skewness	0.26	-0.74	-0.18
Kurtosis	2.81	2.92	2.05

Note: SD – standard deviation, IC – institutional care group, BFC – biological family group.

Table 5

IRT analyses results of the ARFA-RUS Synonyms Subtest

Model	**1-PL**		**2-PL**			**2-PL without item № 18**		
Model fit	AIC = 18644.76, BIC = 18761.44, Log-likelihood = -9296.38		AIC = 18382.94, BIC = 18607.32, Log-likelihood = -9141.47			AIC = 17501.7, BIC = 17717.11, Log-likelihood = -8702.85		
Item no.	***b***	**Item fit**	***a***	***b***	**Item fit**	***a***	***b***	**Item fit**
1	-2.04	*p* = 1.000	0.95	-1.99	*p* = 1.000	0.93	-2.03	*p* = 1.000
2	0.11	*p* = .980	0.83	0.11	*p* = 1.000	0.84	0.11	*p* = 1.000
3	-1.67	***p* = .010**	2.14	-1.05	*p* = 1.000	2.14	-1.05	*p* = 1.000
4	-0.46	*p* = .604	1.00	-0.44	*p* = .941	0.10	-0.44	*p* = .356
5	-1.56	***p* = .010**	1.53	-1.14	*p* = .386	1.49	-1.15	*p* = .198
6	-0.84	*p* = .545	1.17	-0.72	*p* = .891	1.16	-0.72	*p* = 1.000
7	-0.54	*p* = .535	0.97	-0.52	*p* = .277	0.98	-0.52	*p* = .366
8	-1.42	*p* = 1.000	0.71	-1.74	*p* = .975	0.70	-1.77	*p* = 1.000
9	-1.80	***p* = .020**	1.57	-1.29	*p* = 1.000	1.53	-1.30	*p* = .574
10	-0.87	*p* = .100	1.03	-0.81	*p* = .792	1.95	-0.80	*p* = .842
11	-0.02	*p* = .594	0.70	-0.02	***p* = .010**	0.68	-0.02	***p* = .020**
12	-0.44	*p* = .970	0.33	-1.05	*p* = .277	0.31	-1.11	*p = .485*
13	0.11	*p* = .861	0.48	0.21	*p* = .861	0.49	0.21	*p = .931*
14	-1.27	***p* = .010**	2.32	-0.80	*p* = .792	2.32	-0.79	*p = .762*
15	-0.29	*p* = .891	0.53	-0.44	*p* = .584	0.53	-0.44	*p = 1.000*
16	-1.13	***p* = .030**	1.17	-0.97	*p* = .703	1.17	-0.97	*p = .673*
17	-0.23	*p* = .980	0.87	-0.24	*p* = .564	0.88	-0.24	*p = .564*
18	0.03	*p* = .178	0.40	0.08	***p* = .049**	–	–	–
9	-0.55	***p* = .010**	1.94	-0.39	*p* = .941	1.92	-0.39	*p* = .832
20	-1.02	*p* = .119	1.46	-0.78	*p* = 1.000	1.45	-0.78	*p* = .693
21	-1.17	*p* = .753	0.97	-1.12	*p* = .970	0.98	-1.11	*p* = .792
22	-1.55	***p* = .010**	1.31	-1.22	*p* = .406	1.30	-1.23	*p* = .267
23	-2.29	*p* = .861	1.31	-2.34	*p* = .277	0.87	-2.38	***p* = .049**
24	0.50	*p* = .832	0.39	1.07	*p* = .990	0.39	1.07	*p* = .980
25	-1.16	*p* = .891	1.05	-1.06	*p* = 1.000	1.04	-1.06	*p* = .723

Note: *a* – discrimination, *b* – difficulty. AIC – Akaike information criteria, BIC – Bayesian information criterion. Item fit was calculated by Monte Carlo simulation. Poorly fitting items have *p* <.05 (marked in bold).

Table 6

Difficulty and discrimination parameters in the IC and BFC groups

Item no.	BFC		IC		Wald χ^2, *p*-value
	a	*b*	*a*	*b*	
1	1.24	-2.46	0.16	-8.07	10.86, $p = .014$
2	0.66	-0.63	–	–	–
3	2.05	-1.96	–	–	–
4	0.78	-1.28	1.21	-1.44	0.32, $p = .854$
5	1.36	-2.08	–	–	–
6	0.72	-2.06	1.84	-1.63	3.89, $p = .207$
7	0.59	-1.83	1.37	-1.48	2.74, $p = .300$
8	0.61	-3.15	0.06	-10.86	7.70, $p = .046$
9	1.43	-2.23	–	–	–
10	1.13	-1.65	0.33	-1.52	10.31, $p = .015$
11	1.10	-0.46	-0.05	-10.30	37.00, $p < .001$
12	0.58	-0.90	-0.14	-0.39	23.96, $p < .001$
13	0.20	-1.03	0.92	-1.11	4.87, $p = .143$
14	2.04	-1.72	–	–	–
15	0.59	-1.09	0.21	-0.67	3.01, $p = .289$
16	0.99	-1.91	–	–	–
17	0.74	-1.06	–	–	–
19	2.31	-1.01	1.70	-1.33	5.35, $p = .128$
20	1.32	-1.68	–	–	–
21	0.55	-3.30	0.90	-1.91	1.66, $p = .472$
22	1.22	-2.17	–	–	–
23	1.22	-2.67	0.04	-30.31	13.87, $p = .004$
24	0.30	0.59	–	–	–
25	0.89	-2.02	–	–	–

Note: a – discrimination*, b* – difficulty. Items 2, 3, 5, 9, 14, 16, 17, 20, 22, 24, 25 were used as anchors, therefore their parameters are equivalent across groups (marked with en-dashes). Item 18 was eliminated in the previous steps of analysis.

References

1. Andringa S. et al. Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach // Language Learning. 2012. № SUPPL. 2 (62). C. 49–78.
2. Armstrong R. et al. Change in receptive vocabulary from childhood to adulthood: associated mental health, education and employment outcomes // International Journal of Language and Communication Disorders. 2017.
3. Asimina, Ralli M., Melpomeni S., Alexandra T. Language and Psychosocial Skills of Institutionalized Children in Greece // The Open Family Studies Journal. 2017. № 1 (9). C. 76–87.
4. Baker F.B. The basics of item response theory / F.B. Baker, ERIC, 2001. 186 c.
5. Beck I.L., McKeown M.G. Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction // Elementary School Journal. 2007. № 3 (107). C. 251–271.
6. Bent T. et al. Individual differences in the perception of regional, nonnative, and disordered speech varieties // The Journal of the Acoustical Society of America. 2016. № 5 (140). C. 3775–3786.
7. Beverly B.L., McGuinness T.M., Blanton D.J. Communication and academic challenges in early adolescence for children who have been adopted from the former Soviet Union // Language, Speech, and Hearing Services in Schools. 2008. № 3 (39). C. 303–313.
8. Biemiller A., Slonim N. Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition // Journal of Educational Psychology. 2001.
9. Bock R.D., Aitkin M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm // Psychometrika. 1981. № 4 (46). C. 443–459.
10. Carrow-Woolfolk E. Comprehensive Assessment of Spoken Language–Second Edition (CASL-2). Torrance, CA: Western Psychological Services, 2017.
11. Chalmers R.P. Mirt: A multidimensional item response theory package for the R environment // Journal of Statistical Software. 2012. № 6 (48).
12. Chalmers R.P., Counsell A., Flora D.B. It Might Not Make a Big DIF: Improved Differential Test Functioning Statistics That Account for Sampling Variability // Educational and Psychological Measurement. 2016. № 1 (76). C. 114–140.
13. Cunningham A.E., Stanovich K.E. What Reading Does for the Mind // American Educator. 1998. № 1 (22). C. 137–149.
14. Desmarais C. et al. Sentence comprehension in postinstitutionalized school-age children // Journal of Speech, Language, and Hearing Research. 2012. № 1 (55). C. 45–54.
15. Drasgow F., Lissak R.I. Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. // Journal of Applied Psychology. 1983.
16. Dunn D.M. Peabody picture vocabulary test-fifth edition. Bloomington, MN.: Pearson, 2019.
17. Embretson S.E., Reise S.P. Item response theory for psychologists Mahwah, N.J.: Lawrence Erlbaum Associates Publishers, 2000.
18. Fernald A., Marchman V.A., Weisleder A. SES differences in language processing skill and vocabulary are evident at 18 months // Developmental Science. 2013.
19. Fromkin V. et al. The development of language in genie: a case of language acquisition beyond the "critical period" // Brain and Language. 1974. № 1 (1). C. 81–107.
20. Gindis B. Cognitive, Language, and Educational Issues of Children Adopted from Overseas Orphanages // Journal of Cognitive Education and Psychology. 2005. № 3 (4). C. 291–315.
21. Glennen S., Bright B.J. Five years later: Language in school-age internationally adopted children // Seminars in Speech and Language. 2005. T. 26. № 1. 86–101 c.
22. Goulden R., Nation P., Read J. How large can a receptive vocabulary be? // Applied Linguistics. 1990. № 4 (11). C. 341–363.

23. Greenough W.T., Black J.E., Wallace C.S. Experience and brain development. // Child development. 1987. № 3 (58). C. 539–559.
24. Hammill D.D. et al. TOAL-4: Test of Adolescent and Adult Language—Fourth Edition // 2007.
25. Hart B., Risley T.R. Meaningful differences in the everyday experience of young American children. / B. Hart, T.R. Risley, Paul H Brookes Publishing, 1995.
26. Hart B., Risley T.R. The early catastrophe: The 30 million word gap by age 3 // American educator. 2003. № 1 (27). C. 4–9.
27. Hemphill L., Tivnan T. The Importance of Early Vocabulary for Literacy Achievement in High-Poverty Schools // Journal of Education for Students Placed at Risk (JESPAR). 2008. № 4 (13). C. 426–451.
28. Hiebert E.H., Kamil M.L. Teaching and learning vocabulary: Bringing research to practice Routledge, 2005. 1–279 c.
29. Horn J.L. A rationale and test for the number of factors in factor analysis // Psychometrika. 1965. № 2 (30). C. 179–185.
30. Hu L.T., Bentler P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives // Structural Equation Modeling. 1999. № 1 (6). C. 1–55.
31. Ivanova M. et al. AutoRAT at your fingertips: Introducing the new Russian Aphasia Test on a tablet // Frontiers in Psychology. 2016. (7).
32. Janse E., Jesse A. Working memory affects older adults' use of context in spoken-word recognition // Quarterly Journal of Experimental Psychology. 2014. № 9 (67). C. 1842–1862.
33. Johnson M.S. Marginal maximum likelihood estimation of item response models in R // Journal of Statistical Software. 2007. № 10 (20). C. 1–24.
34. Kamata A., Bauer D.J. A note on the relation between factor analytic and item response theory models // Structural Equation Modeling: A Multidisciplinary Journal. 2008. № 1 (15). C. 136–153.
35. Kornilov S.A. et al. Language Outcomes in Adults with a History of Institutionalization: Behavioral and Neurophysiological Characterization // Scientific Reports. 2019. № 1 (9). C. 4252.
36. Langer M.M. et al. Item response theory detected differential item functioning between healthy and ill children in quality-of-life measures // Journal of Clinical Epidemiology. 2008. № 3 (61). C. 268–276.
37. Linden W.J. van der Handbook of item response theory CRC Press, 2016. 1–595 c.
38. Loman M.M. et al. Postinstitutionalized children's development: Growth, cognitive, and language outcomes // Journal of Developmental and Behavioral Pediatrics. 2009. № 5 (30). C. 426–434.
39. Lord F.M. Applications of item response theory to practical testing problems Hillsdale, NJ: Lawrence Erlbaum Associates, 1980. 1–274 c.
40. Magis D. et al. A general framework and an R package for the detection of dichotomous differential item functioning // Behavior Research Methods. 2010. № 3 (42). C. 847–862.
41. Martinková P. et al. Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments // CBE Life Sciences Education. 2017. № 2 (16). C. rm2.
42. Mayberry R.I. Cognitive development in deaf children: The interface of language and perception in neuropsychology // Handbook of Neuropsychology. 2002. № Part II (8). C. 71–107.
43. McDermott J. et al. Psychosocial deprivation, executive functions, and the emergence of socio-emotional behavior problems // Frontiers in Human Neuroscience . 2013. T. 7. 167 c.
44. McGregor K.K. et al. Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth // International Journal of Language and Communication Disorders. 2013. № 3 (48). C. 307–319.
45. Merz E.C., McCall R.B., Wright A.J. Attention and language as mediators of academic outcomes following early psychosocial deprivation // International Journal of Behavioral Development. 2013. № 5 (37). C. 451–459.

46. Milton J., Treffers-Daller J. Vocabulary size revisited: The link between vocabulary size and academic achievement // Applied Linguistics Review. 2013.
47. Muhamedrahimov R.J. et al. Structural characteristics of the institutional environment for young children // Psychology in Russia: State of the Art. 2016.
48. Nation K. Lexical learning and lexical processing in children with developmental language impairments // Philosophical Transactions of the Royal Society B: Biological Sciences. 2014. Т. 369. № 1634. 20120387 c.
49. National Reading Panel Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction // NIH Publication No. 00-4769. 2000.
50. Paek I., Cole K. Using R for Item Response Theory Model Applications / I. Paek, K. Cole, Routledge, 2019.
51. Quinn J.M., Wagner R.K. Using Meta-analytic Structural Equation Modeling to Study Developmental Change in Relations Between Language and Literacy // Child Development. 2018. № 6 (89). С. 1956–1969.
52. R Core Team (2019) R: A language and environment for statistical computing. // Accessed 1st April 2019. 2019.
53. Raju N.S. The area between two item characteristic curves // Psychometrika. 1988. № 4 (53). С. 495–502.
54. Reise S.P., Henson J.M. A discussion of modem versus traditional psychometrics as applied to personality assessment scales // Journal of personality assessment. 2003. № 2 (81). С. 93–103.
55. Richland L.E., Burchinal M.R. Early Executive Function Predicts Reasoning Development // Psychological Science. 2012. № 1 (24). С. 87–92.
56. Rizopoulos D. Itm: An R package for latent variable modeling and item response theory analyses // Journal of Statistical Software. 2006. № 5 (17). С. 1–25.
57. Rodríguez-Aranda C., Jakobsen M. Differential contribution of cognitive and psychomotor functions to the age-related slowing of speech production // Journal of the International Neuropsychological Society. 2011. № 5 (17). С. 807–821.
58. Scott K.A., Roberts J.A., Glennen S. How well do children who are internationally adopted acquire language? a meta-analysis // Journal of Speech, Language, and Hearing Research. 2011. № 4 (54). С. 1153–1169.
59. Shao Z. et al. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults // Frontiers in Psychology. 2014. № JUL (5). С. 772.
60. Stahl S.A., Fairbanks M.M. The Effects of Vocabulary Instruction: A Model-Based Meta-Analysis // Review of Educational Research. 1986.
61. Steinberg L., Thissen D. Item response theory. Oxford library of psychology. / New York, NY, US: Oxford University Press, 2013. 336–373 с.
62. Takane Y., Leeuw J. De On the relationship between item response theory and factor analysis of discretized variables // Psychometrika. 1987. № 3 (52). С. 393–408.
63. Tirella L.G. et al. Time use in Russian Baby Homes // Child: Care, Health and Development. 2008.
64. Unsworth N., Spillers G.J., Brewer G.A. Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance // Quarterly Journal of Experimental Psychology. 2011. № 3 (64). С. 447–466.
65. Wechsler D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). San Antonio, TX: Pearson, 2008.
66. Wechsler D. Wechsler intelligence scale for children–Fifth Edition (WISC-V) Bloomington, MN: Pearson, 2014.

67. Widhiarso W. Haryanta Examining method effect of synonym and antonym test in verbal abilities measure // Europe's Journal of Psychology. 2015.
68. Wiig E.H., Secord W.A., Semel E. Clinical evaluation of language fundamentals: CELF-5 Bloomington, MN: Pearson, 2013.
69. Williams K. Expressive vocabulary test, third edition. Bloomington, MN: Pearson, 2018.
70. Windsor J. et al. Language acquisition with limited input: Romanian institution and foster care // Journal of Speech, Language, and Hearing Research. 2007. № 5 (50). C. 1365–1381.
71. Windsor J. et al. Effect of foster care on young children's language learning // Child Development. 2011. № 4 (82). C. 1040–1046.
72. Windsor J. et al. Effect of foster care on language learning at eight years: Findings from the Bucharest early intervention project // Journal of Child Language. 2013. № 3 (40). C. 605–627.
73. Yap M.J. et al. Individual differences in visual word recognition: Insights from the English Lexicon Project // Journal of Experimental Psychology: Human Perception and Performance. 2012. № 1 (38). C. 53–79.
74. Zanon C. et al. An application of item response theory to psychological test development // Psicologia: Reflexão e Crítica. 2016. № 1 (29). C. 1–10.
75. Prikhoda N.A. Russian Language Development Assessment as a Standardized Technique for Assessing Communicative Function in Children Aged 3–9 Years. Psikhologicheskaya nauka i obrazovanie = Psychological Science and Education, 2016. Vol. 21, no. 3, pp. 25–33. doi:10.17759/pse.2016210304. (In Russ., abstr. in Engl.)

Electronic supplementary material

Below is the link to the electronic supplementary material.

1. https://mfr.osf.io/render?url=https%3A%2F%2Fosf.io%2F93hdx%2Fdownload

Схожи или отличны? Применение Item Response Theory для анализа теста «Синонимы» на выборке взрослых с опытом и без опыта институционализации

Логвиненко Т.И.*
СПбГУ, Санкт-Петербург, Россия,
ORCID: https://orcid.org/0000-0001-7430-1963,
e-mail: logvinenkota.spb@gmail.com

Таланцева О.И.**
СПбГУ, Санкт-Петербург, Россия,
ORCID: https://orcid.org/0000-0002-7555-1216,
e-mail: talantcevaoksana@gmail.com

Волохова Е.М.***
СПбГУ, Санкт-Петербург, Россия,
ORCID: https://orcid.org/0000-0002-1711-0447,
e-mail: human.nerpa@gmail.com

Халаф Ш.****
Университет Хьюстона, Хьюстон, США,
ORCID: https://orcid.org/0000-0002-8809-590X,
e-mail: shiva.khalaf@times.uh.edu

Григоренко Е.Л.*****
СПбГУ, Санкт-Петербург, Россия;
Университет Хьюстона, Хьюстон, США,
ORCID: https://orcid.org/0000-0001-9646-4181, e-mail:
elena.grigorenko@times.uh.edu

В России существует недостаток валидных и стандартизированных инструментов для оценки языкового домена у подростков и взрослых. Для восполнения этого пробела была создана языковая тестовая батарея ARFA-RUS...

Для цитаты:
Логвиненко Т.И., Таланцева О.И., Волохова Е.М., Халаф Ш., Григоренко Е.Л. Схожи или отличны? Применение Item Response Theory для анализа теста «Синонимы» на выборке взрослых с опытом и без опыта институционализации // Моделирование и анализ данных. 2020. Том 10. № 1. С. 35–63. DOI: 10.17759/mda.2020100102 (In Engl., abstr. In Russ.)

Логвиненко Татьяна Игоревна, Инженер-исследователь лаборатории междисциплинарных исследований развития человека, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, ORCID: https://orcid.org/0000-0001-7430-1963, e-mail: logvinenkota.spb@gmail.com

*****Таланцева Оксана Игоревна,*** Инженер-исследователь лаборатории междисциплинарных исследований развития человека, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, ORCID: https://orcid.org/0000-0002-7555-1216, e-mail: talantcevaoksana@gmail.com

Методика использовалась в масштабном исследовательском проекте, посвященном долгосрочным эффектам институционального ухода на развитие. человека. В настоящей работе в качестве первого шага валидизации методики мы изучили психометрических свойств субтеста «Синонимы» батареи ARFA-RUS с помощью подхода item response theory (IRT). Результаты IRT-анализа показали надёжность теста для оценки низких и средних уровней способности, однако тест требуется дополнить более сложными заданиями, чтобы охватить больший диапазон способности. Субтест «Синонимы» батареи ARFA-RUS оказался слабо подходящим для группы взрослых с опытом институционализации: в этой группе вычисленная латентная переменная объясняла значительно меньшую долю дисперсии по сравнению с группой взрослых, воспитанных в биологических семьях. На уровне отдельных заданий два вопроса продемонстрировали парадоксальные закономерности: вероятность правильного ответа уменьшалась при увеличении способности. Кроме того, одно задание было исключено из окончательной версии субтеста «Синонимы» из-за его плохого соответствия модели и низкой дискриминативной способности.

Ключевые слова: Item Response Theory, психометрика, дифференцированное функционирование заданий, языковые способности, синонимы, тестирование.

Финансирование: Исследование выполнено при финансовой поддержке Правительства Российской Федерации, грант № 14.Z50.31.0027 «Влияние ранней депривации на био-поведенческие показатели развития ребенка», ведущий ученый – Григоренко Е.Л.

******Волохова Екатерина Михайловна,*** Инженер-исследователь лаборатории междисциплинарных исследований развития человека, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия, ORCID: https://orcid.org/0000-0002-1711-0447, e-mail: human.nerpa@gmail.com

*******Халаф Шива,*** PhD, научный сотрудник Техасского института измерений, оценки и статистики, Университет Хьюстона, Хьюстон, США, ORCID: https://orcid.org/0000-0002-8809-590X, e-mail: shiva.khalaf@times.uh.edu

********Григоренко Елена Леонидовна,*** PhD, д.п.н., профессор, главный научный сотрудник лаборатории междисциплинарных исследований развития человека, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия; Университет Хьюстона, Хьюстон, США, ORCID: https://orcid.org/0000-0001-9646-4181, e-mail: elena.grigorenko@times.uh.edu

Благодарность

Авторы благодарят Мэй Тэн за помощь в редакции текста публикации.

Литература

1. Andringa S. et al. Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach // Language Learning. 2012. № SUPPL. 2 (62). С. 49–78.
2. Armstrong R. et al. Change in receptive vocabulary from childhood to adulthood: associated mental health, education and employment outcomes // International Journal of Language and Communication Disorders. 2017.
3. Asimina, Ralli M., Melpomeni S., Alexandra T. Language and Psychosocial Skills of Institutionalized Children in Greece // The Open Family Studies Journal. 2017. № 1 (9). С. 76–87.
4. Baker F.B. The basics of item response theory / F.B. Baker, ERIC, 2001. 186 с.
5. Beck I.L., McKeown M.G. Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction // Elementary School Journal. 2007. № 3 (107). С. 251–271.

6. Bent T. et al. Individual differences in the perception of regional, nonnative, and disordered speech varieties // The Journal of the Acoustical Society of America. 2016. № 5 (140). С. 3775–3786.
7. Beverly B.L., McGuinness T.M., Blanton D.J. Communication and academic challenges in early adolescence for children who have been adopted from the former Soviet Union // Language, Speech, and Hearing Services in Schools. 2008. № 3 (39). С. 303–313.
8. Biemiller A., Slonim N. Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition // Journal of Educational Psychology. 2001.
9. Bock R.D., Aitkin M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm // Psychometrika. 1981. № 4 (46). С. 443–459.
10. Carrow-Woolfolk E. Comprehensive Assessment of Spoken Language–Second Edition (CASL-2). Torrance, CA: Western Psychological Services, 2017.
11. Chalmers R.P. Mirt: A multidimensional item response theory package for the R environment // Journal of Statistical Software. 2012. № 6 (48).
12. Chalmers R.P., Counsell A., Flora D.B. It Might Not Make a Big DIF: Improved Differential Test Functioning Statistics That Account for Sampling Variability // Educational and Psychological Measurement. 2016. № 1 (76). С. 114–140.
13. Cunningham A.E., Stanovich K.E. What Reading Does for the Mind // American Educator. 1998. № 1 (22). С. 137–149.
14. Desmarais C. et al. Sentence comprehension in postinstitutionalized school-age children // Journal of Speech, Language, and Hearing Research. 2012. № 1 (55). С. 45–54.
15. Drasgow F., Lissak R.I. Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. // Journal of Applied Psychology. 1983.
16. Dunn D.M. Peabody picture vocabulary test-fifth edition. Bloomington, MN.: Pearson, 2019.
17. Embretson S.E., Reise S.P. Item response theory for psychologists Mahwah, N.J.: Lawrence Erlbaum Associates Publishers, 2000.
18. Fernald A., Marchman V.A., Weisleder A. SES differences in language processing skill and vocabulary are evident at 18 months // Developmental Science. 2013.
19. Fromkin V. et al. The development of language in genie: a case of language acquisition beyond the “critical period” // Brain and Language. 1974. № 1 (1). С. 81–107.
20. Gindis B. Cognitive, Language, and Educational Issues of Children Adopted from Overseas Orphanages // Journal of Cognitive Education and Psychology. 2005. № 3 (4). С. 291–315.
21. Glennen S., Bright B.J. Five years later: Language in school-age internationally adopted children // Seminars in Speech and Language. 2005. Т. 26. № 1. 86–101 с.
22. Goulden R., Nation P., Read J. How large can a receptive vocabulary be? // Applied Linguistics. 1990. № 4 (11). С. 341–363.
23. Greenough W.T., Black J.E., Wallace C.S. Experience and brain development. // Child development. 1987. № 3 (58). С. 539–559.
24. Hammill D.D. et al. TOAL-4: Test of Adolescent and Adult Language—Fourth Edition // 2007.
25. Hart B., Risley T.R. Meaningful differences in the everyday experience of young American children. / B. Hart, T.R. Risley, Paul H Brookes Publishing, 1995.
26. Hart B., Risley T.R. The early catastrophe: The 30 million word gap by age 3 // American educator. 2003. № 1 (27). С. 4–9.
27. Hemphill L., Tivnan T. The Importance of Early Vocabulary for Literacy Achievement in High-Poverty Schools // Journal of Education for Students Placed at Risk (JESPAR). 2008. № 4 (13). С. 426–451.
28. Hiebert E.H., Kamil M.L. Teaching and learning vocabulary: Bringing research to practice Routledge, 2005. 1–279 с.

29. Horn J.L. A rationale and test for the number of factors in factor analysis // Psychometrika. 1965. № 2 (30). C. 179–185.
30. Hu L.T., Bentler P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives // Structural Equation Modeling. 1999. № 1 (6). C. 1–55.
31. Ivanova M. et al. AutoRAT at your fingertips: Introducing the new Russian Aphasia Test on a tablet // Frontiers in Psychology. 2016. (7).
32. Janse E., Jesse A. Working memory affects older adults' use of context in spoken-word recognition // Quarterly Journal of Experimental Psychology. 2014. № 9 (67). C. 1842–1862.
33. Johnson M.S. Marginal maximum likelihood estimation of item response models in R // Journal of Statistical Software. 2007. № 10 (20). C. 1–24.
34. Kamata A., Bauer D.J. A note on the relation between factor analytic and item response theory models // Structural Equation Modeling: A Multidisciplinary Journal. 2008. № 1 (15). C. 136–153.
35. Kornilov S.A. et al. Language Outcomes in Adults with a History of Institutionalization: Behavioral and Neurophysiological Characterization // Scientific Reports. 2019. № 1 (9). C. 4252.
36. Langer M.M. et al. Item response theory detected differential item functioning between healthy and ill children in quality-of-life measures // Journal of Clinical Epidemiology. 2008. № 3 (61). C. 268–276.
37. Linden W.J. van der Handbook of item response theory CRC Press, 2016. 1–595 c.
38. Loman M.M. et al. Postinstitutionalized children's development: Growth, cognitive, and language outcomes // Journal of Developmental and Behavioral Pediatrics. 2009. № 5 (30). C. 426–434.
39. Lord F.M. Applications of item response theory to practical testing problems Hillsdale, NJ: Lawrence Erlbaum Associates, 1980. 1–274 c.
40. Magis D. et al. A general framework and an R package for the detection of dichotomous differential item functioning // Behavior Research Methods. 2010. № 3 (42). C. 847–862.
41. Martinková P. et al. Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments // CBE Life Sciences Education. 2017. № 2 (16). C. rm2.
42. Mayberry R.I. Cognitive development in deaf children: The interface of language and perception in neuropsychology // Handbook of Neuropsychology. 2002. № Part II (8). C. 71–107.
43. McDermott J. et al. Psychosocial deprivation, executive functions, and the emergence of socio-emotional behavior problems // Frontiers in Human Neuroscience. 2013. T. 7. 167 c.
44. McGregor K.K. et al. Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth // International Journal of Language and Communication Disorders. 2013. № 3 (48). C. 307–319.
45. Merz E.C., McCall R.B., Wright A.J. Attention and language as mediators of academic outcomes following early psychosocial deprivation // International Journal of Behavioral Development. 2013. № 5 (37). C. 451–459.
46. Milton J., Treffers-Daller J. Vocabulary size revisited: The link between vocabulary size and academic achievement // Applied Linguistics Review. 2013.
47. Muhamedrahimov R.J. et al. Structural characteristics of the institutional environment for young children // Psychology in Russia: State of the Art. 2016.
48. Nation K. Lexical learning and lexical processing in children with developmental language impairments // Philosophical Transactions of the Royal Society B: Biological Sciences. 2014. T. 369. № 1634. 20120387 c.
49. National Reading Panel Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction // NIH Publication No. 00-4769. 2000.
50. Paek I., Cole K. Using R for Item Response Theory Model Applications / I. Paek, K. Cole, Routledge, 2019.

51. Quinn J.M., Wagner R.K. Using Meta-analytic Structural Equation Modeling to Study Developmental Change in Relations Between Language and Literacy // Child Development. 2018. № 6 (89). С. 1956–1969.
52. R Core Team (2019) R: A language and environment for statistical computing. // Accessed 1st April 2019. 2019.
53. Raju N.S. The area between two item characteristic curves // Psychometrika. 1988. № 4 (53). С. 495–502.
54. Reise S.P., Henson J.M. A discussion of modem versus traditional psychometrics as applied to personality assessment scales // Journal of personality assessment. 2003. № 2 (81). С. 93–103.
55. Richland L.E., Burchinal M.R. Early Executive Function Predicts Reasoning Development // Psychological Science. 2012. № 1 (24). С. 87–92.
56. Rizopoulos D. Itm: An R package for latent variable modeling and item response theory analyses // Journal of Statistical Software. 2006. № 5 (17). С. 1–25.
57. Rodríguez-Aranda C., Jakobsen M. Differential contribution of cognitive and psychomotor functions to the age-related slowing of speech production // Journal of the International Neuropsychological Society. 2011. № 5 (17). С. 807–821.
58. Scott K.A., Roberts J.A., Glennen S. How well do children who are internationally adopted acquire language? a meta-analysis // Journal of Speech, Language, and Hearing Research. 2011. № 4 (54). С. 1153–1169.
59. Shao Z. et al. What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults // Frontiers in Psychology. 2014. № JUL (5). С. 772.
60. Stahl S.A., Fairbanks M.M. The Effects of Vocabulary Instruction: A Model-Based Meta-Analysis // Review of Educational Research. 1986.
61. Steinberg L., Thissen D. Item response theory. Oxford library of psychology. / New York, NY, US: Oxford University Press, 2013. 336–373 с.
62. Takane Y., Leeuw J. De On the relationship between item response theory and factor analysis of discretized variables // Psychometrika. 1987. № 3 (52). С. 393–408.
63. Tirella L.G. et al. Time use in Russian Baby Homes // Child: Care, Health and Development. 2008.
64. Unsworth N., Spillers G.J., Brewer G.A. Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance // Quarterly Journal of Experimental Psychology. 2011. № 3 (64). С. 447–466.
65. Wechsler D. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). San Antonio, TX: Pearson, 2008.
66. Wechsler D. Wechsler intelligence scale for children–Fifth Edition (WISC-V) Bloomington, MN: Pearson, 2014.
67. Widhiarso W., Haryanta Examining method effect of synonym and antonym test in verbal abilities measure // Europe's Journal of Psychology. 2015.
68. Wiig E.H., Secord W.A., Semel E. Clinical evaluation of language fundamentals: CELF-5 Bloomington, MN: Pearson, 2013.
69. Williams K. Expressive vocabulary test, third edition. Bloomington, MN: Pearson, 2018.
70. Windsor J. et al. Language acquisition with limited input: Romanian institution and foster care // Journal of Speech, Language, and Hearing Research. 2007. № 5 (50). С. 1365–1381.
71. Windsor J. et al. Effect of foster care on young children's language learning // Child Development. 2011. № 4 (82). С. 1040–1046.
72. Windsor J. et al. Effect of foster care on language learning at eight years: Findings from the Bucharest early intervention project // Journal of Child Language. 2013. № 3 (40). С. 605–627.

73. Yap M.J. et al. Individual differences in visual word recognition: Insights from the English Lexicon Project // Journal of Experimental Psychology: Human Perception and Performance. 2012. № 1 (38). С. 53–79.
74. Zanon C. et al. An application of item response theory to psychological test development // Psicologia: Reflexão e Crítica. 2016. № 1 (29). С. 1–10.
75. Прихода Н.А. Оценка развития русского языка (ОРРЯ) как стандартизованная методика диагностики коммуникативной функции у детей от 3 до 9 лет // Психологическая наука и образование. 2016. № 3 (21). С. 25–33.

Дополнительный электронный материал

Ниже приведена ссылка на дополнительный электронный материал.

1. https://mfr.osf.io/render?url=https%3A%2F%2Fosf.io%2F93hdx%2Fdownload