Clinical Psychology and Special Education
2022, vol. 11, no. 2, pp. 138–157.
DOI: 10.17759/cpse.2022110209
ISSN: 2304-0394 (online)

Клиническая и специальная психология
2022. Том 11. № 2. С. 138–157.
DOI: 10.17759/cpse.2022110209
ISSN: 2304-0394 (online)

A Russian Translation of the BRIEF2 Disproportionately Flags Typical Russian and Previously Institutionalized Individuals on Validity Scales

Lisa K. Chinn
University of Houston, Houston, TX, USA,
ORCID: https://orcid.org/0000-0001-8498-9909, e-mail: lisa.chinn@times.uh.edu

Darya A. Momotenko
Sirius University of Science and Technology, Federal territory "Sirius", Russia,
ORCID: https://orcid.org/0000-0003-2544-5420, e-mail: daryamomotenko@gmail.com

Elena L. Grigorenko
University of Houston, Houston, TX, USA,
ORCID: https://orcid.org/0000-0001-9646-4181, e-mail: elena.grigorenko@times.uh.edu

The Behavior Rating Inventory of Executive Function (BRIEF) is a commonly used tool for researchers and clinicians to assess executive functioning, especially in individuals with learning or other developmental disorders. Although it has been translated and used in multiple countries, the BRIEF has only been officially normed by its manufacturers in U.S. samples. In order to further the ideal of cultural sensitivity in psychological testing and examine whether the BRIEF functions appropriately in Russia and in its distinct subpopulations (e.g., individuals with an early history of adversity), we assessed the performance of its built-in validity scales by administering the BRIEF2 Self-Report Form to a Russian sample (n=572) either raised in biological families (n=315) or with a history of institutionalized care (n=257). Results indicate that, compared to U.S. norms, a large number of this sample was flagged for inconsistent or abnormal answers on the BRIEF2 validity scales. This finding highlights the importance of validity checks when psychological tools are used in new cultures, languages, and samples. Results point to a need for fine-tuning of the BRIEF2 Self-Report Form and/or its scoring system before widespread adaptation among Russian clinicians and researchers.

Keywords: BRIEF, behavior rating inventory of executive function, institutionalization, executive functioning, scale, cross-cultural, validation.

Funding. This research was supported in part by grant no. 14.Z50.31.0027 from the Government of the Russian Federation (P.I.: Elena L. Grigorenko) and by the Hugh Roy and Lillie Cranz Cullen Distinguished Professor Chair of the University of Houston (to Elena L. Grigorenko).

Chinn L.K., Momotenko D.A., Grigorenko E.L.
A Russian Translation of the BRIEF2
Disproportionately Flags Typical Russian and
Previously Institutionalized Individuals on Validity Scales
Clinical Psychology and Special Education
2022, vol. 11, no. 2, pp. 138–157.

Чинн Л.К., Момотенко Д.А., Григоренко Е.Л.
Русский перевод BRIEF2 непропорционально
отмечает русскоговорящих людей с опытом
и без опыта институционализации
по шкалам валидности
Клиническая и специальная психология
2022. Том 11. № 2. С. 138–157.

Acknowledgements. We would like to thank Marina Zhukova for her feedback, and Daria Kostina, Julia Nedoshivina, and Anastasia Sukmanova for data cleaning, management, and preparation.

For citation: Chinn L.K., Momotenko D.A., Grigorenko E.L. A Russian Translation of the BRIEF2 Disproportionately Flags Typical Russian and Previously Institutionalized Individuals on Validity Scales. *Klinicheskaia i spetsial'naia psikhologiia=Clinical Psychology and Special Education*, 2022. Vol. 11, no. 2, pp. 138–157. DOI: 10.17759/cpse.2022110209

Русский перевод BRIEF2 непропорционально отмечает русскоговорящих людей с опытом и без опыта институционализации по шкалам валидности

Чинн Л.К.
Хьюстонский университет, Хьюстон, Техас, США,
ORCID: https://orcid.org/0000-0001-8498-9909, e-mail: lisa.chinn@times.uh.edu

Момотенко Д.А.
Научно-технологический университет «Сириус», ФТ «Сириус», Россия,
ORCID: https://orcid.org/0000-0003-2544-5420, e-mail: daryamomotenko@gmail.com

Григоренко Е.Л.
Хьюстонский университет, Хьюстон, Техас, США,
ORCID: https://orcid.org/0000-0001-9646-4181, e-mail: elena.grigorenko@times.uh.edu

Опросник The Behavior Rating Inventory of Executive Function (BRIEF) является широко используемым инструментом для исследователей и клиницистов для оценки исполнительных функций, особенно у людей с нарушениями обучения или другими нарушениями развития. Несмотря на то, что BRIEF был переведен и используется во многих странах, он был официально валидизирован только создателями теста на выборках испытуемых из США. Чтобы расширить культурное разнообразие в психологическом тестировании и проверить, адекватно ли функционирует BRIEF в России и в отдельных подгруппах (например, у людей, имеющих опыт переживания неблагоприятных событий на ранних этапах жизни), мы оценили эффективность

встроенных в опросник шкал валидности, предложив заполнить форму самоотчета BRIEF2 русскоязычной выборке (N=572), представители которой либо выросли в биологических семьях (n=315), либо в прошлом были институционализированы (n=257). Результаты показывают, что по сравнению с американскими нормами большое количество ответов в выборке данного исследования были помечены как непоследовательные или нетипичные по шкалам валидности BRIEF2. Такой результат подчеркивает важность проверки валидности в случаях, когда психологические инструменты применяются в новых культурах, языках и выборках. Кроме того, результаты указывают на необходимость доработки формы самоотчета BRIEF2 и/или системы начисления баллов перед проведением широкомасштабной адаптации среди российских клиницистов и исследователей.

Ключевые слова: BRIEF, behavior rating inventory of executive function, институционализация, исполнительные функции, шкала, кросс-культурный, валидация.

Финансирование. Это исследование было частично поддержано грантом № 14.Z50.31.0027 Правительства Российской Федерации (руководитель — Елена Л. Григоренко) и заслуженным профессором Хью Рой и Лилли Кранц Каллен Хьюстонского университета (Елене Л. Григоренко).

Благодарности. Авторы благодарят Марину Жукову за обратную связь, а также Дарью Костину, Юлию Недошивину и Анастасию Сукманову за подготовку данных.

Для цитаты: *Чинн Л.К., Момотенко Д.А., Григоренко Е.Л.* Русский перевод BRIEF2 непропорционально отмечает русскоговорящих людей с опытом и без опыта институционализации по шкалам валидности [Электронный ресурс] // Клиническая и специальная психология. 2022. Том 11. № 2. С. 138–157. DOI: 10.17759/cpse.2022110209

The Behavior Rating Inventory of Executive Function (BRIEF) [32] is an executive functioning scale widely used among clinical psychologists. The BRIEF is recommended for use in psychological disorders, such as attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD), and has also been found to show changes in a variety of medical conditions, including traumatic brain injury [8] and Alzheimer's disease [29]. Although it has been featured in over 400 peer-reviewed papers and used in multiple countries [31], to our knowledge, the BRIEF has not yet been validated in Russia in general nor in previously institutionalized Russian adolescents and adults specifically. Furthermore, the producers of the version of the BRIEF studied here — the BRIEF2 for ages 5–18 — state that it has only been officially normed within English-speaking/reading samples and standardized based on USA census statistics [16]. However, translated versions exist and have been studied outside the USA [15]. The current study assessed the performance of the BRIEF2 Self-Report Form validity scales when the form was translated to Russian and administered to Russian adolescents and adults who were either previously institutionalized or raised in biological families.

Many efforts to examine the effectiveness of the BRIEF in languages other than English and countries other than the USA have involved analyzing its factor structure. For example, the parent and teacher forms of a French version of the BRIEF were found to be reliable and to have a good model fit for two- or three-factor models comprised of the individual subscales [15]. A Dutch translation of the BRIEF was characterized by high internal consistency, high test-retest stability, and a factor structure similar to U.S. participants [19]. Some articles on BRIEF translations do not mention the validity scales at all, and others do not analyze their performance in depth. For example, an article on the performance and factor structure of the BRIEF in a Spanish clinical sample mentioned excluding participants who were flagged by the Negativity or Inconsistency scales but did not report the percentage of participants excluded or compare performance on these scales to USA norms [13]. In the current study, we checked the performance of the validity scales, even though this method is not common in the literature on BRIEF translations. This decision was because the validity scales would potentially be used in future work to exclude participants who answer atypically, and we wanted to check whether such exclusions could be made using U.S. thresholds.

Extensive testing of translated measures is also important because cultural bias in psychological testing is a major ethical concern for clinicians [30] and researchers. For instance, mean scores on assessments such as IQ tests tend to differ among minority groups [30], making the cultural validity of psychological assessments a topic of heated debate. Besides linguistic equivalence in translations, other factors can vary across cultures, such as equivalence in constructs measured, familiarity with the type of assessment, and cultural relevance of measures [6]. A majority of psychological research has been conducted in Western Industrialized Rich Democracies; therefore, the validity of assessments in diverse cultures is a concern [23]. Additionally, many cross-cultural instrument adaptation studies have relied on factor analysis and underutilized other strategies [2]. To further the goal of testing assessment instruments thoroughly before using them in new cultures, we aimed to test the BRIEF2 Self Report Form in Russian and previously institutionalized samples, starting here with its validity scales.

According to the BRIEF2 manual, rater characteristics such as parent education level and race/ethnicity did not contribute meaningfully to BRIEF self-report standardized scores [16]. For example, parent education accounted for less than 3% of the variability in the self-report data, and race/ethnicity was not significantly related to BRIEF2 self-report scores. Therefore, some factors that may vary cross-culturally, such as education and race/ethnicity, might not contribute substantially to differences in BRIEF performance. Nonetheless, it is important to test performance of the BRIEF before using it extensively in new languages and countries.

The standardization sample used for the BRIEF2 manual included participants with no history of special education, psychotropic medication usage, or neurological disorders (such as ADHD or ASD), with 803 participants who completed the self-report from [16]. The current study, which was part of a larger project on institutionalization, did not screen for type of education or medication use, so our samples might not match the standardization samples on those aspects. Unfortunately, the BRIEF2 has not been standardized in

a previously institutionalized US sample, which would have been interesting to compare to our Russian institutionalized sample. The manual includes general clinical samples, as well as samples prone to deficits in executive functioning, including ADHD and ASD.

Institutional care, defined here as care in government institutions without a family structure (e.g., orphanages or baby homes), in Russia, is often characterized by psychosocial deprivation, frequent changes in caregivers, and the deprivation of close individual contact between the child and caregiver [22; 34]. Children in institutions may have psychosocial difficulties associated with the absence of personal space due to living in the same room as multiple other children, a low level of adaptive care, and stigmatization from peers with whom they attend school [27; 35]. Russian institutions disproportionately contain children with disabilities, although typically developing infants and children are also placed in institutions [20]. U.S. institutions are often termed "group homes," which house between 7 and 12 children, or "residential care." These institutions primarily house those who need services such as therapy and medical care for severe behavioral issues or mental disorders [10; 38] but also contain typical children awaiting foster care placement. In both the U.S. and Russia, institutions are highly structured [22; 34; 10]. Children in U.S. and Russian institutional care may face some of the same struggles, such as trauma from changes in guardians [27; 35]. One difference is that although institutional care is improving in quality and becoming less common in Russia, it is still more common than in the U.S. For instance, approximately 19% of Russian children without parental care are placed in institutions [28], versus around 10% of children without parental care in the U.S. [11]. This 19% comprises a large number of individuals because the rate of public care is high in the Russian Federation (1673 in 100,000 children as of 2021) [28].

Executive Function Assessment

Executive function assessments are important for clinicians because executive functioning is linked to academic achievement [1], health-related quality of life [5], language ability [17], and other major life outcomes. Executive functioning is also implicated in a range of disorders, including ADHD [24], depression [37], and schizophrenia [9], among many others, as well as early life experiences. Specifically, a history of institutional care has been shown to be related to deficits in executive function, as measured with cognitive performance tasks and neuroimaging [21; 25; 26]. The number of individuals with a history of such care is large (see above); therefore, executive functioning assessments for clinicians working with previously institutionalized individuals are important.

The BRIEF2 is one such assessment that has practical advantages over some executive function assessments commonly used in research or clinical practice. It contains seven subscales of executive functioning in just one 55-item questionnaire: Inhibit, Self-Monitoring, Shifting, Emotional Control, Task Completion, Working Memory, and Planning and Organization [16]. Therefore, it is comprehensive and quick. It does not require expensive neuroimaging or computing equipment. It also does not involve multiple tasks with different sets of instructions to assess different aspects of executive functioning, which might be difficult for those with attentional deficits to complete. If demonstrated to be valid in Russia and in previously institutionalized samples, it would be a useful tool for clinical

evaluation and research in Russia in general and in particular for studying individuals with a history of institutionalization. The current study examined the validity scales built into the BRIEF2 Self-Report form designed to detect atypical, inconsistent, or overly negative responses.

Method

Participants. Recruitment took place in major Russian cities. Those at least 18 years of age gave their written informed consent on a consent form approved by the Ethical Committee of the St. Petersburg State University #02-199 on May 3, 2017. Those under 18 had their caregivers sign consent forms. Participants were compensated with 1000 rubles, which came out to an hourly rate approximately equal to the average local hourly wage at the start of the study in 2017. They were primarily recruited through orphanages, social assistance centers, and secondary educational institutions (e.g., lyceums, technical schools, colleges), and some were self-referred via Internet ads. Participants were included in the institutional care (IC) group if institutional records or the participant reported at least 6 months of institutionalization on the initial study screening. The biological family care (BFC) group was raised exclusively in their biological families. These participants were recruited to fall within a similar age range and educational level as the IC group. Participants were native Russian speakers.

Participants were adolescents and adults who took part in a larger project on institutionalization outcomes (n=677). Of these, 654 completed the BRIEF2 Self Report Form, and 636 completed a medical questionnaire. After excluding four participants who selected multiple answers on some BRIEF items and six who failed to complete all BRIEF items, there were 625 participants who completed both the BRIEF2 Self Report Form and the medical questionnaire. Of the 625 who completed both the BRIEF2 and the medical questionnaire, 53 were excluded because they did not select “no” on medical questions asking if they had a recent history of head trauma or neurological illness. The final sample contained 572 participants (331 female, 241 male; 315 BFC, 257 IC). Of these participants, 182 were adolescents (68 BFC, 114 IC; 103 female, 79 male; ages 15–17 years, mean age=16.38, SD=0.64), and 390 were adults (247 BFC, 143 IC; 228 female, 162 male; ages 18–38 years, mean age=22.47 years, SD=4.68). In this sample, 550 participants completed the Culture Fair IQ Test (CFIT) [7]. Because IQ was not the primary focus of the current study or an exclusion criterion, we did not exclude participants who did not complete the CFIT. Participants were involved in a larger project that included additional assessments not analyzed here, including 4 EEG tasks, a handedness questionnaire, and a behavioral battery of language ability.

Because participants were recruited primarily through educational institutions with the goal of approximately matching the IC and BFC groups on age and education, we did not control our sample to make it perfectly representative of the overall Russian population. Median income for adults in our sample was greater than 30000 rubles, versus a median of 28345 rubles in the Russian Federation at the time of data collection in 2017 [12] (https://eng.rosstat.gov.ru/). In our sample, 25% of participants aged 15–29 had a job. Based on Russian government statistics stating that 68.5% of individuals aged 15–72 have jobs and that only 20.2% of the employed were 15–29 years, we estimate that

approximately 13% of the 15- to 29-year-old general population was employed at the time of data collection [12]. In our sample, employment rates may have been higher because we focused recruitment on secondary educational institutions with a high percentage of students from orphanages (such as technical schools). Students from these types of institutions may have been more likely to have additional earnings in comparison with full-time students of bachelor's degree-granting universities or individuals with less education. The IC and BFC groups were not matched on all income-related variables. For example, satisfaction with income was lower in the IC group ($\chi^2_{(1)}$=30.945, p<.001), as was employment ($\chi^2_{(1)}$=37.127, p<.001).

Assessments

Culture Fair IQ Test (CFIT). The Culture Fair Intelligence Test (CFIT; Scale 2; Form B [7]) was used to assess non-verbal intelligence (IQ). IQ data from this study did not fit a normal distribution and, therefore, could not be used to calculate standardized scores. Instead, IQ scores were calculated using the Cattell Culture Fair IQ Key standard scores for Form A, Scale 2 based on both a USA sample and a UK sample. The USA scoring most closely gave the data a normal distribution, so we used the USA scoring guide.

BRIEF2 Self-Report Form. The BRIEF2 Self-Report Form is a 55-item questionnaire originally designed for ages 11–18. It takes approximately 10 minutes to complete. Each item on the BRIEF2 describes a behavior that represents a problem with executive functioning and asks the participant to rate whether they never, sometimes, or often have the problem [16]. The item scores are then summed into composite scores (1 — never, 2 — sometimes, 3 — often) within each of 7 subscales (Working memory, Inhibit, Self-Monitor, Shift, Emotional Control, Task Completion, Plan/Organize). Therefore, higher scores on the BRIEF2 indicate worse executive functioning. For the current study, all items were translated to Russian and then translated a second time into English (a commonly accepted method called back-translation [4]) to check the accuracy of the first translation. At the time of the current analyses, accuracy of the Russian translations was checked once again by a native Russian speaker with English fluency. To be able to analyze adult and adolescent data together in analyses that are part of our larger project on institutionalization, we used the same version of the questionnaire (the BRIEF2 Self-Report) for all participants (adults and adolescents), and in the current paper checked whether the validity scales performed differently in the two age groups. Many items on the BRIEF2 for ages 11–18 and the Adult (BRIEF-A [33]) Self-Report forms are identical or very similar, suggesting that the 11- to 18-year-old version might also accurately assess young adults.

Validity Scales. The BRIEF2 Self-Report contains three validity indicators designed to flag individuals with questionable responses: Infrequency, Inconsistency, and Negativity scales.

Infrequency. The Infrequency scale contains three questions that are not part of the executive function subscales and are highly unusual to endorse, even for severely cognitively impaired participants according to the professional manual for the BRIEF2 (e.g., endorsing that they forget their own name) [16]. Thus, these items are designed to detect highly atypical answers and may reflect falsehoods or extreme impairments.

Inconsistency. The Inconsistency scale identifies inconsistency between answers to similar questions. Discrepancy scores are computed for pairs of similar items to check for response inconsistencies.

Negativity. The Negativity scale contains items that are part of the executive functioning scales, and it identifies when a participant gives an abnormally large number of "often" responses on negative items (e.g., endorsing that the participant often talks at the wrong time).

Results

Demographic Analysis. Although the BRIEF2 manual explains that validity measures have been tested in a typical standardization sample as well as a variety of clinical samples, including cognitive impairment, we checked IQ scores on the CFIT for the current sample to make sure that the majority of participants did not have very low IQ. The mean IQ score was 99.31 ($Mean_{BFC}$=105.67, SD_{BFC}=13.15; $Mean_{IC}$=91.14, SD_{IC}=12.80; see Figure 1). Thirteen participants had scores lower than 70 (2.3%; 5 with 57, 3 with 62, 5 with 66), which is a common threshold for diagnosing intellectual disability [14]. They were retained in the current analyses due to their small number and the BRIEF manufacturer's suggestion that the validity scales work in most clinical groups and that the BRIEF works with a broad range of participants. Furthermore, only a small number of participants flagged on the validity scales had IQ below 70, so excluding them would not have changed the results by much (see Table 1 and 2 captions and Negativity Scale results below).

BRIEF Reliability*.* Cronbach's alpha values were computed for our sample to compare to the BRIEF2 manual [16]. Scores for our whole dataset (alpha range 0.72 to 0.83 across scales) and for the individual BFC (0.67 to 0.83) and IC (0.75 to 0.82) groups largely overlapped with the standardization (0.81 to 0.90) and atypical clinical (0.71 to 0.85) samples described in the BRIEF2 manual [16].

CFIT Reliability. Reliability of the CFIT in our overall sample was checked using Cronbach's alpha. For our whole sample, the range of Cronbach's alpha across subtests was 0.88 to 0.91. Alpha values for the BFC (0.77 to 0.90) and IC (0.81 to 0.87) groups largely overlapped with each other. The CFIT manual does not clarify how reliability was computed for the manual, so we cannot directly compare the manual to our sample, but the manual gives a value of 0.76 for consistency over items [7].

Infrequency Scale. According to the professional manual for the BRIEF2 [16], questionable scores on the infrequency scale, indicated by selecting "sometimes" or "often" on at least one infrequency item, indicate a greater than 99th percentile score, even in most clinical groups and in individuals with cognitive impairment (tested in U.S. samples). However, in the current overall sample, questionable scores on this scale occurred for 8.6% of participants (see Table 1), which is significantly more often than 1% ($z=18.27$, $p<.0001$). Even for the current study subgroup with the lowest rate of questionable scores (BFC adolescents; 4.4%; see Figure 2), the rate was significantly higher than the 1% indicated by U.S. norms ($z=2.82$, $p<.01$).

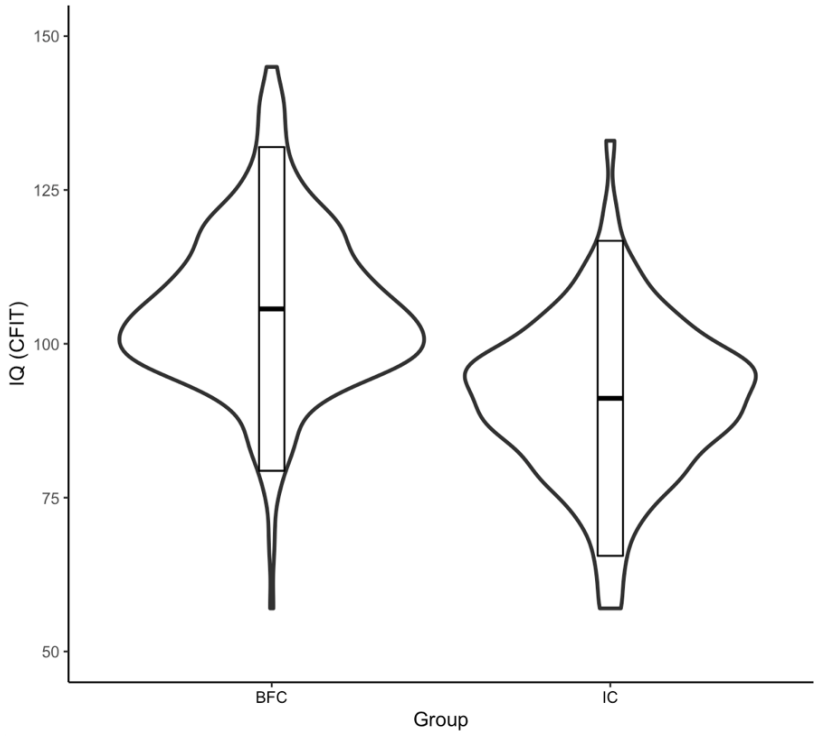

Figure 1. Violin plots of IQ scores by institutionalization status. Rectangles represent two standard deviations from the mean

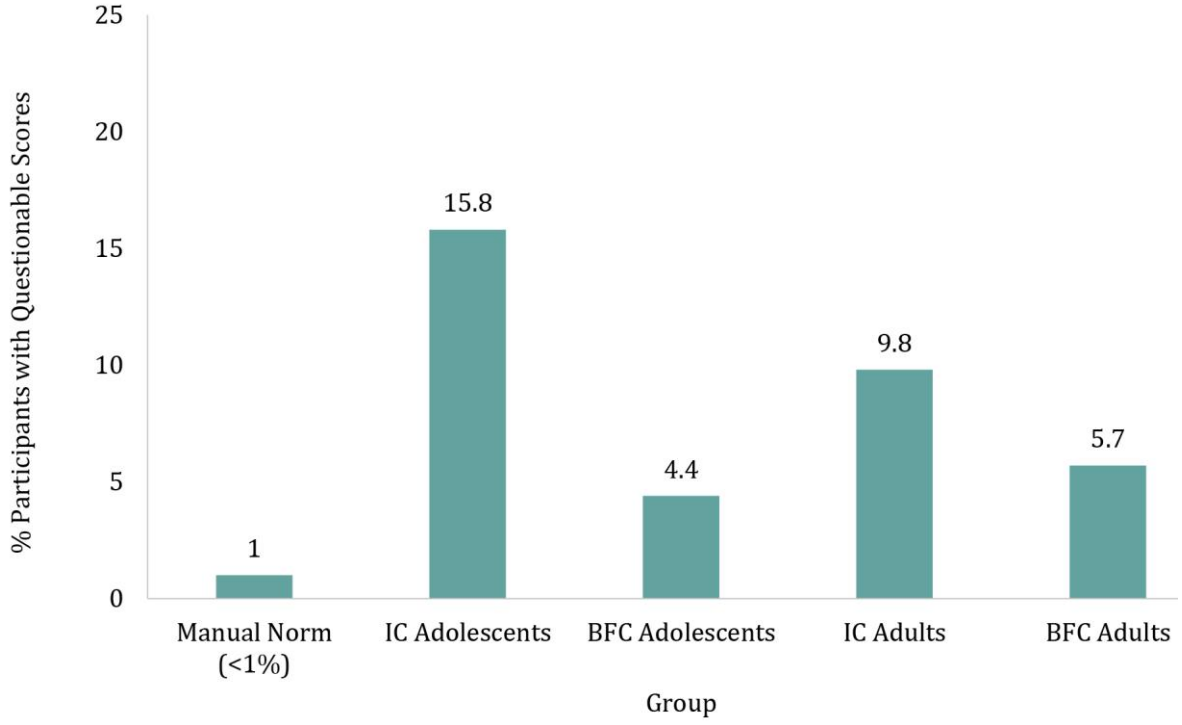

Figure 2. Percentages of questionable scores on infrequency items split by age groups and institutionalization history

Table 1

Count data and percentages of questionable scores on infrequency items split by age groups and institutionalization history

	Questionable Scores
Total	49/572=8.6%*
Adolescents	21/182=11.5%
IC Adolescents	18/114=15.8%
BFC Adolescents	3/68=4.4%
Adults	28/390=7.2%
IC Adults	14/143=9.8%
BFC Adults	14/247=5.7%
IC Total	32/257=12.5%
BFC Total	17/315=5.4%

Notes. * — with individuals with IQ below 70 excluded, 44 participants had questionable scores on the infrequency scale. Therefore, a high proportion of low IQ individuals were flagged on this scale (4 out of 13). However, even with those who have an IQ below 70 excluded, the percentage of participants flagged on this scale in this sample remains high compared to a U.S. sample (8.2% versus <1%).

Inconsistency Scale*.* Inconsistency scores are computed by summing the absolute value of the differences between similar items, and the questionnaire contains 8 similar item pairs. A score less than or equal to 5 is acceptable (98th percentile or lower according to U.S. norms), 6-7 is questionable (99th percentile), and 8+ indicates inconsistent responses (>99th percentile). In the current sample, both questionable + inconsistent scores (101/572 or 17.7%) and inconsistent scores alone (14/572 or 2.4%) occurred significantly more often than 1% of the time ($z=40.04$, $p<.0001$; $z=3.48$, $p<.001$, respectively; see Table 2 and Figure 3), which is the maximum amount of questionable and inconsistent scores according to the manual normed on U.S. participants.

Negativity Scale*.* To compute negativity scores, the number of items with a score of "often" on 8 negativity items is counted. A total of 6 or less is considered acceptable, 7 is elevated, and 8 is highly elevated. According to the normative samples used in the BRIEF2 manual, elevated scores are considered 99th percentile, and highly elevated are above the 99th percentile (according to U.S. samples). Only one participant in the current sample scored above a 6 (see Table 3), which was significantly less than the 1% indicated as likely based on U.S. norms ($z=-1.98$, $p<.05$). The one flagged individual had an IQ above 70, so exclusions based on IQ would not have changed the results significantly.

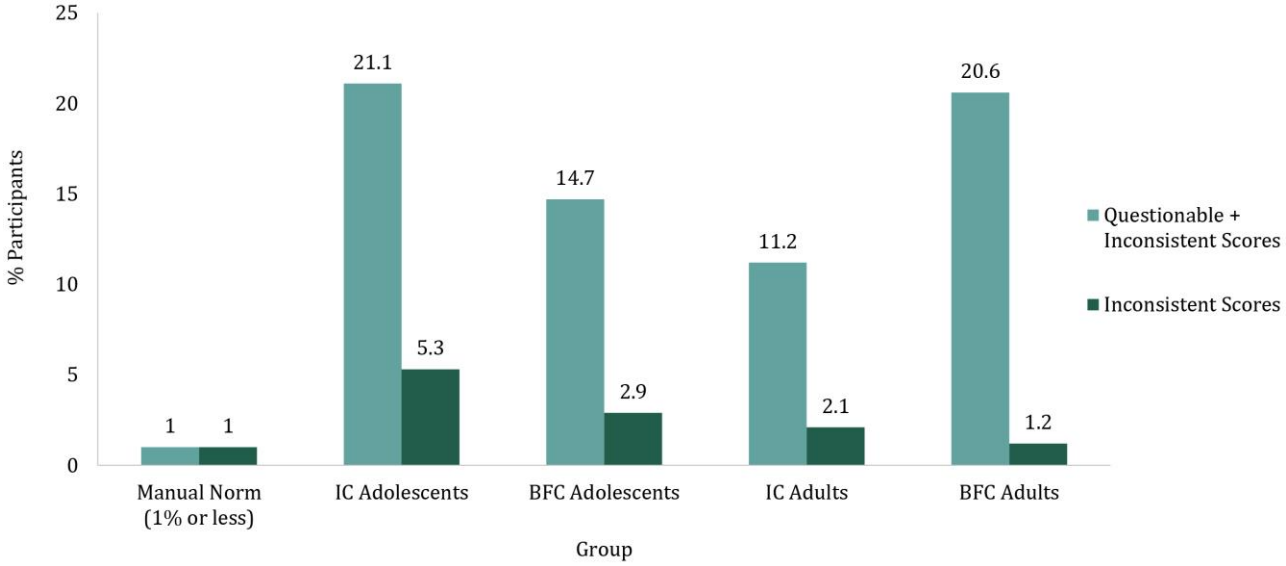

Figure 3. Percentages of questionable + inconsistent or inconsistent only scores derived from inconsistency items, split by age groups and institutionalization history

Table 2

Count data and percentages of questionable + inconsistent or inconsistent only scores derived from inconsistency items, split by age groups and institutionalization history

	Questionable + Inconsistent Scores	Inconsistent Scores
Total	101/572=17.7%*	14/572=2.4%*
Adolescents	34/182=18.7%	8/182=4.4%
IC Adolescents	24/114=21.1%	6/114=5.3%
BFC Adolescents	10/68=14.7%	2/68=2.9%
Adults	67/390=17.2%	6/390=1.5%
IC Adults	16/143=11.2%	3/143=2.1%
BFC Adults	51/247=20.6%	3/247=1.2%
IC Total	40/257=15.7%	9/257=3.5%
BFC Total	61/315=19.4%	5/315=1.6 %

Notes. * — with individuals with IQ below 70 excluded, 94 participants had questionable or inconsistent scores and 12 had inconsistent scores on the inconsistency scale. Therefore, a high proportion of low IQ individuals were flagged on this scale (7 out of 13). However, even with those who have an IQ below 70 excluded, the percentage of participants flagged on this scale in this sample remains high compared to a U.S. sample (17.5% questionable or inconsistent versus 1%; 2.2% inconsistent versus <1%).

Table 3

Count data and percentages of elevated or highly elevated + inconsistent scores on negativity items, split by age groups and institutionalization history

	Scores
Total Elevated	0/572=0%
Total Highly Elevated	1/572=0.2%
Adolescents Highly Elevated	1/182=0.5%
IC Adolescents	1/114=0.9%
BFC Adolescents	0/68=0%
Adults Highly Elevated	0/390=0%
IC Adults	0/143=0%
BFC Adults	0/247=0%
IC Total Highly Elevated	1/257=0.4%
BFC Total Highly Elevated	0/315=0%

Relations between Validity Scales, Covariates, and BRIEF. Because percentages of individuals with questionable scores on the Infrequency and Inconsistency scales were higher than indicated in the BRIEF2 Professional Handbook with norms created within the United States, despite excluding those with head trauma or neurological illness and despite not explicitly recruiting a clinical population, further exploration was warranted. The negativity scale is not analyzed in this section because only one participant had an elevated negativity score.

Infrequency and Inconsistency. If the reason for the high number of questionable responses on the infrequency scale was that some participants selected answers arbitrarily or inaccurately, it seemed likely that participants with questionable responses on the infrequency scale participants would also be prone to answering similar questions inconsistently and be flagged by the inconsistency scale. Although a majority of participants with questionable scores on the infrequency scale had acceptable scores on the inconsistency scale (34/49; see Figure 4), a higher proportion of individuals with questionable infrequency scores had questionable or inconsistent answers on the inconsistency scale than those with acceptable infrequency scores ($\chi^2_{(1,572)}=6.19$, $p<.05$).

IQ and Validity Scales. For the Infrequency and Inconsistency scales, a logistic generalized linear model testing the effects of IQ, Group (IC, BFC), Age Group (Adult,

Adolescent), and the IQ x Group interaction on whether scores were questionable was run using the glm function in R with a binomial distribution specified.

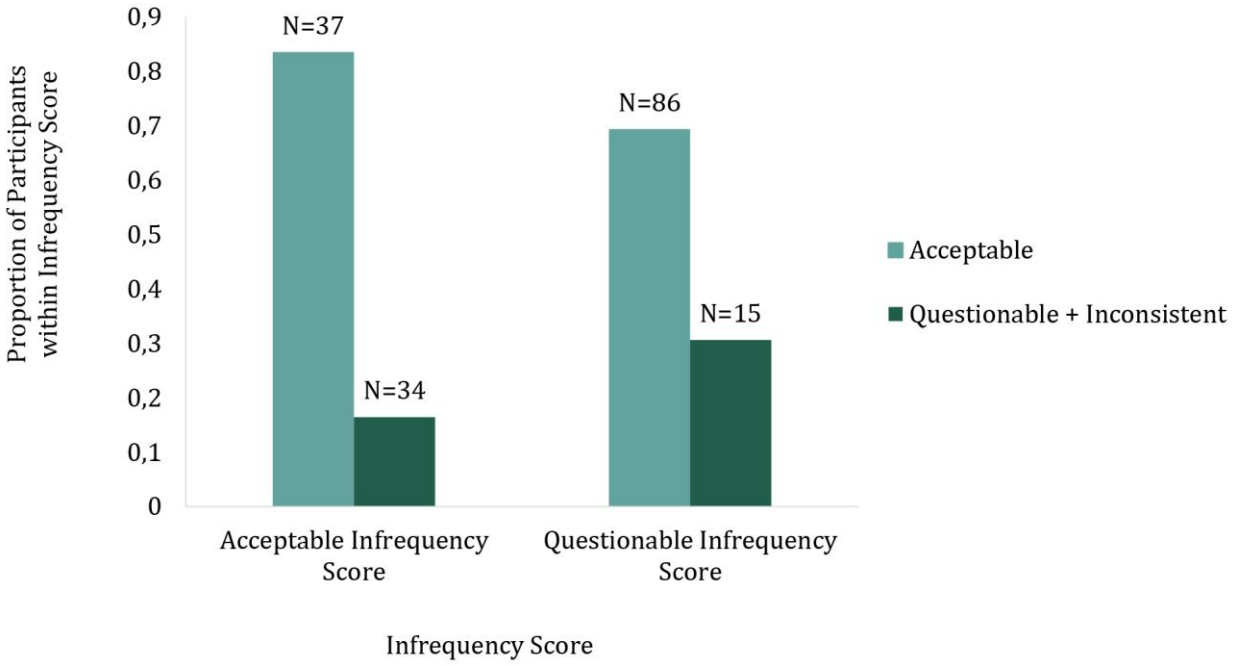

Figure 4. Proportions of participants with acceptable or questionable infrequency scores and those with acceptable or questionable + inconsistent inconsistency scores, plus count data

Infrequency. The IQ x Group interaction had a statistically significant effect on the probability of getting a questionable infrequency score ($\chi^2_{(1)}$=7.62, p<.01; see Figure 5). Follow-up tests indicated that in the BFC group, IQ scores were not significantly related to the probability of getting a questionable infrequency score (z=1.70, p=.09), and in the IC group, IQ scores were negatively related to the probability of getting a questionable infrequency score (z=-2.30, p<.05). The main effect of Institutionalization was statistically significant ($\chi^2_{(1)}$=9.14, p<.01), such that those with a history of institutionalization were more likely to have questionable infrequency answers than those raised in biological families. The Age effect was not statistically significant.

Inconsistency. Questionable and inconsistent scores were combined into one "non-acceptable" category. There were no statistically significant effects of IQ, Group (BFC, IC), IQ x Group, or Age Group (adolescent, adult) on the probability of getting non-acceptable answers on the inconsistency scale.

BRIEF Performance and Validity Scales. A linear regression tested the effects of validity (Acceptable or Questionable) on each BRIEF subscale, controlling for Group (IC, BFC), Age Group (Adolescent, Adult), and Gender (male, female). For all subscales of the BRIEF and for both Infrequency and Inconsistency subscales, questionable answers were associated with worse executive function (see Table 4).

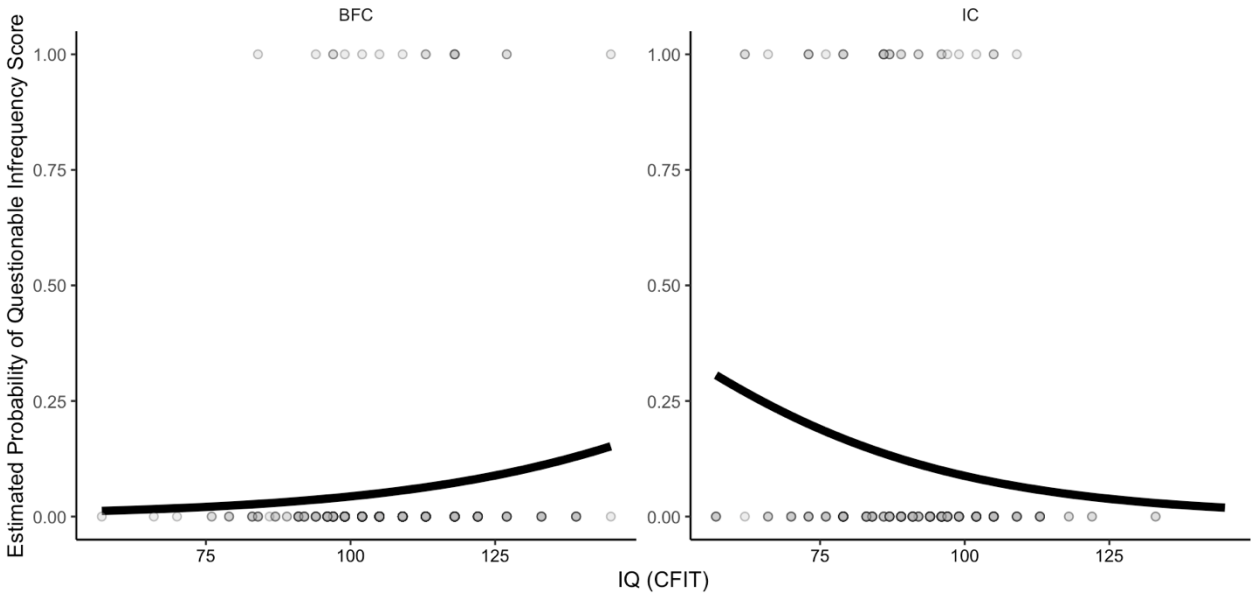

Figure 5. IQ by predicted infrequency probability in each group

Notes. Black lines represent the GLM-predicted probability of having a questionable infrequency score. Dots represent the raw data, with 0=acceptable and 1=questionable. Overlapping dots are darker.

Table 4

F-values, p-values, and directions of effects for linear models testing effects of questionable Infrequency and Inconsistency scores on each BRIEF subscale

BRIEF Subscale	Infrequency F-value	Infrequency p-value	Inconsistency F-value	Inconsistency p-value
Inhibit	17.19	<.0001	17.68	<.0001
Working Memory	33.11	<.0001	42.88	<.0001
Shift	9.45	<.01	18.17	<.0001
Plan	20.48	<.0001	25.63	<.0001
Self-Monitor	31.74	<.0001	18.06	<.0001
Task Completion	12.91	<.001	48.59	<.0001
Emotional Control	6.60	<.05	22.07	<.0001

Discussion

The current study compared Russian samples raised in exclusively in biological families or at least partially in institutionalized care to U.S. norms on three validity scales

built into the BRIEF2 Self-Report form [16]. Results indicate that for scales designed to flag highly infrequent (abnormal) or inconsistent answers, significantly more individuals were flagged in our overall sample (8.6 & 17.7%) compared to typical and clinical samples in the U.S. (~1% or less). The infrequency scale, designed to detect highly abnormal answers, was also sensitive to IQ and a history of institutionalization. Questionable answers on infrequency and inconsistency scales were also associated with worse executive functioning on all subscales of the BRIEF. Although the BRIEF2 was designed for use in individuals aged 11–18, performance on the infrequent and inconsistent validity scales did not significantly vary by age group (adolescent, adult). Results point to possible cultural differences in responding between the U.S. and Russia and highlight the need for culturally sensitive validity checks before translated measures are used extensively in research or clinical practice.

The high rates of questionable scores on the infrequency and inconsistency subscales and low rate of elevated scores on the negativity scale in the current study were not driven by any particular question(s). For example, the three infrequency items had similar numbers of questionable responses (26, 22, & 24 questionable responses per item). Additionally, even though our sample included previously institutionalized individuals who may be more prone to lower IQ [36] and executive functioning [21; 25; 26] than non-institutionalized individuals, the results seem unlikely to be driven by low functioning in this sample. One reason for this conclusion is that the BRIEF2 validity scales were tested in the U.S. in clinical samples and those with cognitive impairment, and they still showed low levels of questionable answers. Additionally, mean IQ in this sample (99.31) was close to the U.S. average (100). Furthermore, even when considering only the intended age group (our adolescent group) and participants raised in biological families, the Russian sample had more questionable answers on the infrequency scale and inconsistency scales than U.S. norms. Lastly, although several of our 13 participants with IQ below 70 were flagged on the inconsistency and infrequency scales, this was a small percentage of our total participants. Thus, excluding them would still have left us with higher numbers of questionable responses on the infrequency and inconsistency scales. The questionable scores, therefore, do not seem to be driven solely by impaired intelligence (although questionable scores in the IC group were associated with lower IQ) and are likely to be elevated, at least in part, for another reason.

The relatively high number of individuals flagged on infrequency and inconsistency validity scales could be due in part to cross-cultural differences in assessments, specifically unfamiliarity with the type of testing [6]. Although the Russian education system uses multiple choice tests, psychological screening in Russia has historically been less common compared to western countries [3; 18]. Therefore, participants may not have been as comfortable with self-reflection and self-report questionnaires. The differences found here may also be due to unidentified cultural differences between Russia and the U.S.

Another possibility is that the high infrequency and inconsistency scores might reflect a subset of individuals who answered quickly or carelessly. Although participants got an hour break halfway through testing, the larger study visit took approximately five hours. Therefore, some participants may have answered indiscriminately to rush through testing or made errors if they became fatigued during testing. Furthermore, we saw strong

relations between two of the validity scales (inconsistency and infrequency) and the BRIEF subscales that measure executive functioning, such that those with lower executive functioning were more likely to be flagged on the validity scales. (Although note that it remains unclear how much the BRIEF executive function scores can be trusted in individuals flagged by the validity scales.) Careless answering or attentional difficulties could explain the high percentage of participants flagged for inconsistent answers. Additionally, since only one response of "sometimes" or "often" on any of the three infrequency items was required for a participant to be flagged on the infrequency scale, indiscriminate answering would likely flag more than the U.S. norm of 1% of participants on this scale. However, most participants who were flagged on one of these validity scales were not flagged on the other (only 15/572 were flagged on both), suggesting that a majority of questionable scores may not have been driven by one set of participants choosing answers at random.

Low engagement on the BRIEF in some participants could also partially explain why the negativity scale flagged no participants in the current study as "elevated" and just one (~0.1%) as "highly elevated". In order to be flagged on the negativity scale, a participant had to answer "often" on 7-8 out of 8 items on the negativity scale [16]. If a subset of participants was indiscriminately providing a range of answers, they would be unlikely to answer "often" on so many negativity items and could slightly lower the negativity rate. However, it remains unclear why such a small number (1/572) was flagged on this subscale in our sample or how meaningfully different this is from the 1% in U.S. norms, given that reaching 1% would have required only 4-5 more participants to be flagged in our sample.

A limitation of the current study is that we cannot conclusively determine the reason(s) for differences in validity between our sample and U.S. norms, so cultural differences would need to be explored in future research. Additionally, due to a focus on recruiting individuals with a history of institutionalization, our sample did not match the larger population of the Russian Federation on all demographic measures. An additional future direction could be to norm validity scales in larger Russian samples of neurotypical and clinical participants, more similar to the U.S. BRIEF2 scoring manual [16]. Future work could apply the same exclusion criteria as U.S. standardization samples, such as psychotropic medication use. Researchers could also directly compare Russian and U.S. samples in the same study and could try a study design with a smaller number of additional measures to reduce the potential for rushing or fatigue effects. The reason(s) that we saw a Group x IQ interaction on the infrequency scale also remains unclear and should be explored further.

Conclusions

The primary goal of this report was to evaluate the usability of the BRIEF2 Self-Report Form scales in a Russian sample. Cultural validity is important when a scale is used in new populations [6; 23; 30], and here we demonstrated that the BRIEF2 validity scales may not perform the same in Russian and U.S. samples. The infrequency scale also does not perform the same in individuals raised in biological families versus those raised at least partially in institutions. Future work will evaluate the BRIEF2 further, using analyses such as item response theory, confirmatory factor analysis [15], and checking for correlations between

the BRIEF2 scores and neural measures of executive functioning using EEG. Until results are clearer, researchers and clinicians should use translations of this scale in Russian samples with caution.

References

1. Ahmed S.F., Tang S., Waters N.E. et al. Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, 2019, vol. 111, no. 3, pp. 446–458. DOI: 10.1037/EDU0000296

2. Arafat S., Chowdhury H., Qusar M. et al. Cross cultural adaptation & psychometric validation of research instruments: A methodological review. *Journal of Behavioral Health,* 2016, vol. 5, no. 3, pp. 129–136. DOI: 10.5455/JBH.20160615121755

3. Balachova T.N., Levy S., Isurina G.L. et al. Medical psychology in Russia. *Journal of Clinical Psychology in Medical Settings,* 2001, vol. 8, no. 1, pp. 61–68.

4. Beaton D.E., Bombardier C., Guillemin F. et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, 2000, vol. 25, no. 24, pp. 3186–3191. DOI: 10.1097/00007632-200012150-00014

5. Brown T.E., Landgraf J.M. Improvements in executive function correlate with enhanced performance and functioning and health-related quality of life: Evidence from 2 large, double-blind, randomized, placebo-controlled trials in ADHD. *Postgraduate Medical Journal,* 2015, vol. 122, no. 5, pp. 42–51. DOI: 10.3810/PGM.2010.09.2200

6. Byrne B.M. Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation,* 2016, vol. 5, no. 1, pp. 51–65. DOI: 10.1037/IPP0000042

7. Cattell R.B., Cattell A.K.S. Measuring intelligence with the culture fair tests. Champaign, Illinios: Institute for Personality and Ability Testing, 1973, pp. 5–52.

8. Chapman L.A., Wade S.L., Walz N.C. et al. Clinically significant behavior problems during the initial 18 months following early childhood traumatic brain injury. *Rehabilitation Psychology,* 2010, vol. 55, no. 1, pp. 48. DOI: 10.1037/A0018418

9. Chey J., Lee J., Kim Y.S. et al. Spatial working memory span, delayed response and executive function in schizophrenia. *Psychiatry Research,* 2002, vol. 110, no. 3., pp. 259–271. DOI: 10.1016/S0165-1781(02)00105-1

10. Child Welfare Information Gateway. *Group and residential care.* U.S. Department of Health and Human Services. N.d. URL: https://www.childwelfare.gov/topics/outofhome/group-residential-care/ (Accessed: 01.05.2022).

11. Children's Bureau. *The AFCARS report. Adoption and Foster Care Analysis Reporting System.* U.S. Department of Health and Human Services, 2020, vol. 27. URL: https://www.acf.hhs.gov/sites/default/files/documents/cb/afcarsreport27.pdf (Accessed: 01.05.2022).

12. Federal State Statistics Service. N.d. URL: https://eng.rosstat.gov.ru/(Accessed: 30.04.2022).

13. Fernández T.G., González-Pienda J.A., Pérez C.R. et al. Psychometric characteristics of the BRIEF scale for the assessment of executive functions in Spanish clinical population. *Psicothema*, 2014, vol. 26, no. 1, pp. 47–54. DOI: 10.7334/PSICOTHEMA2013.149

14. Fernell E., Gillberg C. Borderline intellectual functioning. *Handbook of Clinical Neurology*, 2020, vol. 174, pp. 77–81. DOI: 10.1016/B978-0-444-64148-9.00006-5

15. Fournet N., Roulin J.L., Monnier C. et al. Multigroup confirmatory factor analysis and structural invariance with age of the Behavior Rating Inventory of Executive Function (BRIEF)-French version. *Child Neuropsychology,* 2015, vol. 21, no. 3, pp. 379–398. DOI: 10.1080/09297049.2014.906569

16. Gioia G.A., Isquith P.K., Guy S.C. et al. BRIEF2 Behavior Rating Inventory of Executive Function. Lutz, FL: Psychological Assessment Resources, Inc., 2015. 334 p.

17. Gooch D., Thompson P., Nash H.M. et al. The development of executive function and language skills in the early school years. *Journal of Child Psychology and Psychiatry*, 2016, vol. 57, no. 2, pp. 180–187. DOI: 10.1111/JCPP.12458

18. Grigorenko E.L. Russian "Defectology": Anticipating Perestroika in the Field. *Journal of Learning Disabilities,* 1998, vol. 31, no. 2, pp. 193–207.

19. Huizinga M., Smidts D.P. Age-related changes in executive function: A normative study with the Dutch version of the Behavior Rating Inventory of Executive Function (BRIEF). *Child Neuropsychology,* 2011, vol. 17, no. 1, pp. 51–66. DOI: 10.1080/09297049.2010.509715

20. Human Rights Watch. Abandoned by the state: Violence, neglect, and isolation for children with disabilities in Russian orphanages. UNICEF, 2014. URL: https://gdc.unicef.org/resource/abandoned-state-violence-neglect-and-isolation-children-disabilities-russian-orphanages (Accessed: 01.05.2022).

21. Lamm C., Troller-Renfree S.V., Zeanah C.H. et al. Impact of early institutionalization on attention mechanisms underlying the inhibition of a planned action. *Neuropsychologia*, 2018, vol. 117, pp. 339–346. DOI: 10.1016/j.neuropsychologia.2018.06.008

22. Lantrip C., Isquith P.K., Koven N.S. et al. Executive function and emotion regulation strategy use in adolescents. *Applied Neuropsychology: Child,* 2016, vol. 5, no. 1, pp. 50–55. DOI: 10.1080/21622965.2014.960567

23. Leong F.T.L., Priscilla Lui P., Kalibatseva Z. Multicultural issues in clinical psychological assessment. *Cambridge Handbook of Clinical Assessment and Diagnosis*, 2019, pp. 25–37. DOI: 10.1017/9781108235433.003

24. Martel M., Nikolas M., Nigg J.T. Executive function in adolescents with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry,* 2007, vol. 46, no. 11, pp. 1437–1444. DOI: 10.1097/CHI.0B013E31814CF953

25. McDermott J.M., Westerlund A., Zeanah C.H. et al. Early adversity and neural correlates of executive function: Implications for academic adjustment. *Developmental Cognitive Neuroscience,* 2012, vol. 2, suppl. 1, pp. S59–S66. DOI: 10.1016/j.dcn.2011.09.008

26. Merz E.C., Harlé K.M., Noble K.G. et al. Executive function in previously institutionalized children. *Child Development Perspectives,* 2016, vol. 10, no. 2, pp. 105–110. DOI: 10.1111/cdep.12170

27. Muhamedrahimov R.J., Grigorenko E.L. Seeing the trees within the forest: Addressing the needs of children without parental care in the Russian Federation. *New Directions for Child and Adolescent Development*, 2015, vol. 147, pp. 101–108. DOI: 10.1002/cad.20080

28. Posarac A., Andreeva E., Bychkov D. et al. Organization and delivery of child protection services in Russia: With two case studies — the Leningrad oblast and the Republic of Tatarstan. Washington, DC: World Bank, 2021. License: CC BY 3.0 IGO. URL: https://openknowledge.worldbank.org/handle/10986/35622 (Accessed: 11.06.2022)

29. Rabin L.A., Roth R.M., Isquith P.K. et al. Self- and informant reports of executive function on the BRIEF-A in MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology,* 2006, vol. 21, no. 7, pp. 721–732. DOI: 10.1016/J.ACN. 2006.08.004

30. Reynolds C.R., Suzuki L.A. Bias in Psychological Assessment. *Handbook of Psychology,* 2nd ed., 2012, vol. 10, pp. 82–112, DOI: 10.1002/9781118133880.HOP210004

31. Roth R.M., Erdodi L.A., McCulloch L.J. et al. Much ado about norming: the Behavior Rating Inventory of Executive Function. *Child Neuropsychology,* 2015, vol. 21, no. 2, pp. 225–233. DOI: 10.1080/09297049.2014.897318

32. Roth R.M., Isquith P.K., Gioia G.A. Assessment of executive functioning using the behavior rating inventory of executive function (BRIEF). In: *Handbook of Executive Functioning*. Springer New York, 2014, pp. 301–331. DOI: 10.1007/978-1-4614-8106-5_18

33. Roth R.M., Isquith P.K., Gioia G.A. BRIEF-A Behavior Rating Inventory of Executive Function — Adult Version. Psychological Assessment Resources, Inc., 2005. URL: https://paa.com.au/product/brief-a/ (Accessed: 30.04.2022).

34. Tirella L.G., Chan W., Cermak S.A. et al. Time use in Russian baby homes. *Child: Care, Health and Development.* 2008, vol. 34, no. 1, pp. 77–86. DOI: 10.1111/j.1365-2214.2007.00766.x

35. Tsvetkova L.A., Grigorenko E.L., Muhamedrahimov R.J. et al. Structural characteristics of the institutional environment for young children. *Psychology in Russia*, 2016, vol. 9, no. 3, pp. 103–112. DOI: 10.11621/pir.2016.0307

36. van IJzendoorn M.H., Luijk P.C.M., Juffer F. IQ of Children Growing Up in Children's Homes: A Meta-Analysis on IQ Delays in Orphanages. *Merrill-Palmer Quarterly,* 2008, vol. 54, no. 3, pp. 341–366. DOI: 10.1353/mpq.0.0002

37. Watkins E., Brown R.G. Rumination and executive function in depression: an experimental study. *Journal of Neurology, Neurosurgery and Psychiatry*, 2002, vol. 72, no. 3, pp. 400–402. DOI: 10.1136/JNNP.72.3.400

38. Wiltz T. Giving group homes a 21st century makeover. The Pew Charitable Trusts, 2018. URL: https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2018/06/14/giving-group-homes-a-21st-century-makeover (Accessed: 01.05.2022).

Information about the authors

Lisa K. Chinn, PhD (Psychology), Postdoctoral Fellow, Texas Institute for Measurement, Evaluation, and Statistics, Department of Psychology, University of Houston, Houston, TX, USA, ORCID: https://orcid.org/0000-0001-8498-9909, e-mail: lisa.chinn@times.uh.edu

Darya A. Momotenko, M.S. (Psychology), PhD Student, Saint-Petersburg State University, Saint-Petersburg, Russia; Junior Researcher, Center for Cognitive Sciences, Sirius University of Science and Technology, Federal territory "Sirius", Russia, ORCID: https://orcid.org/0000-0003-2544-5420, e-mail: daryamomotenko@gmail.com

Elena L. Grigorenko, PhD (Psychology, Behavioral Genetics), Hugh Roy and Lillie Cranz Cullen Distinguished Professor of Psychology, University of Houston, Houston, TX, USA; Adjunct Senior Research Scientist, Moscow State University for Psychology and Education, Moscow, Russia; Professor and Acting Director, Center for Cognitive Sciences, Sirius University of Science and Technology, Federal territory "Sirius", Russia; Research Certified Professor, Baylor College of Medicine, Houston, TX, USA; Adjunct Professor, Child Study Center and Adjunct Senior Research Scientist, Haskins Laboratories, Yale University, New Haven, CT, USA, ORCID: https://orcid.org/0000-0001-9646-4181, e-mail: elena.grigorenko@times.uh.edu

Информация об авторах

Чинн Лиза К., PhD (психология), постдокторант, Техасский институт измерений, оценки и статистики, департамент психологии, Хьюстонский университет, Хьюстон, Техас, США, ORCID: https://orcid.org/0000-0001-8498-9909, e-mail: lisa.chinn@times.uh.edu

Момотенко Дарья Анатольевна, магистр (психология), аспирант, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия; младший научный сотрудник, Центр когнитивных исследований, Научно-технологический университет «Сириус», ФТ «Сириус», Россия, ORCID: https://orcid.org/0000-0003-2544-5420, e-mail: daryamomotenko@gmail.com

Григоренко Елена Леонидовна, PhD (психология, поведенческая генетика), заслуженный профессор психологии Хью Рой и Лилли Кранц Каллен, Хьюстонский университет, Хьюстон, Техас, США; ведущий научный сотрудник, Московский государственный психолого-педагогический университет, Москва, Россия; профессор и и.о. директора, Научный центр когнитивных исследований, Научно-технологический университет «Сириус», ФТ «Сириус», Россия; профессор, Медицинский колледж Бейлора, Хьюстон, Техас, США; адъюнкт-профессор, Центр исследований детства и Лаборатории Хаскинса, Йельский университет, Нью-Хейвен, Коннектикут, США, ORCID: https://orcid.org/0000-0001-9646-4181, e-mail: elena.grigorenko@times.uh.edu

Received: 24.10.2021

Accepted: 06.05.2022

Получена: 24.10.2021

Принята в печать: 06.05.2022