

methods

Adaptation of the Aphasia Bedside Check for Russian

Olga Buivolova

Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia; Research Center of Medical Rehabilitation, Federal Center for Brain Research and Neurotechnologies, Moscow, Russia

Roelien Bastiaanse

Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia; Center for Language and Cognition Groningen (CLCG), University of Groningen, Groningen, The Netherlands

Olga Dragoy

Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia; Research Center of Medical Rehabilitation, Federal Center for Brain Research and Neurotechnologies, Moscow, Russia

Oxana Vinter

Department of Neurology, Moscow City Clinical Hospital No. 31, Moscow, Russia

Victoriia Pozdniakova

Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia

Anastasiia Samoukina

Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia

Anastasia Shlyakhova

Center for Language and Brain, National Research University Higher School of Economics, Moscow, Russia

Evy Visch-Brink

Department of Neurosurgery, Erasmus MC, Rotterdam, the Netherlands

Abstract. The Aphasia Bedside Check for Russian (ABC-Ru) is a screening test that allows the medical staff of a neurological ward to detect speech/language disorders in the first days post-stroke onset. To evaluate whether this test follows modern psychometric standards, we performed two studies. In Study 1, we reported the results of the standardization of the ABC-Ru in a clinical group of people with chronic speech/language disorders ($N = 80$) and a cohort of neurologically healthy individuals ($N = 120$). In Study 2, we validated the results in a group of people in the acute post-stroke period ($N = 20$) with and without speech/language disorders. According to the results of the study, the ABC-Ru can be considered as a valid instrument and can potentially be used in neurological departments for the screening of speech and language disorders.

Correspondence: Olga Buivolova, obuivolova@hse.ru, Center for Language and Brain, National Research University Higher School of Economics, 21/4 Staraya Basmanaya Street, Office 510, Moscow 105066, Russian Federation; Roelien Bastiaanse, ybastiaanse@hse.ru; Olga Dragoy, odragoy@hse.ru; Oxana Vinter, oksvinter@mail.ru; Victoriia Pozdniakova, vpozdneyakova@hse.ru; Anastasiia Samoukina, nastya.samoukina@gmail.com; Anastasia Shlyakhova, aseshly32@gmail.com; Evy Visch-Brink, e.visch-brink@erasmusmc.nl.

Keywords: screening test, standardized assessment, acute stroke, aphasia, Russian

Copyright © 2020. Olga Buivolova, Roelien Bastiaanse, Olga Dragoy, Oxana Vinter, Victoriia Pozdniakova, Anastasiia Samoukina, Anastasia Shlyakhova, Evy Visch-Brink. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original authors are credited and that the original publication in this journal is cited, in accordance with accepted academic practice.

Acknowledgements. We are grateful to Roman A. Cheremin and Galina E. Ivanova for their help with organization of the study and for providing access to people admitted to the Center of Speech Pathology and Neurorehabilitation and Moscow Municipal Clinical Hospital No31. We acknowledge all colleagues from the Center for Language and Brain of the National Research University Higher School of Economics for their comments on the manuscript. We would also like to express our

gratitude to all participants of the present study. The study was supported by the Center for Language and Brain NRU Higher School of Economics, RF Government grant, ag. № 14.641.31.0004.

Received July 31, 2020, accepted September 25, 2020.

Introduction

Stroke is the second most common cause of death in Russia, and eighth most common cause of disability (Institute for Health Metrics and Evaluation, 2018). One of the possible consequences of a stroke, along with motor disabilities, is aphasia. Aphasia is a language impairment that leads to problems with the production and comprehension of language, which seriously affects a person's quality of life (Lucot et al., 2013). To reduce the negative effects of impairment and to predict recovery, it is important to identify speech/language disorders as early as possible.

Neurologists and other medical staff of neurological wards use speech and language screening tests (National Stroke Foundation, 2010) to evaluate the linguistic abilities of the patient at admission to the hospital. Screening is an important step to further diagnostics and for development of a rehabilitation program. In the acute period, speech and language are usually assessed with general screening scales, such as the National Institutes of Health Stroke Scale (NIHSS, National Institute of Neurological Disorders and Stroke [U.S.], 2011), or with specific instruments developed for speech and language screening which are widely used in different languages (for a review, see El Hachoui et al., 2017; Rohde et al., 2018).

In Russian clinical practice, the NIHSS is recommended for screening the patient's state during the first days post-stroke. This scale allows clinicians to assess different functions, including speech and language. However, the NIHSS is not meant for detailed speech and language assessment. The *Aphasia Rapid Test* (Azuar, Leger, Arbizu, & Samson, 2013; Russian version: Buivolova, Vinter, Bastiaanse & Dragoy, 2020) is a standardized screening test for speech and language disorders inspired by the NIHSS. Both scales are suitable for initial screening and they are easy to administer and perform; however, they do not provide information about the core of any language deficit.

Another screening test which is in use in clinics is the "Screening method for examination of a patient with aphasia" (Scherbakova & Kotov, 2015). This test has been developed for identifying language disorders and it allows a neurologist, speech/language therapist, or neuropsychologist to identify the aphasia type according to Luria's classification (Akhutina, 2016). Performing this test takes little time (5–30 min). The instrument is suitable for screening and meets the needs of specialists. Even so, the authors do not provide information about standardization of the test, whether it was normed on a cohort of healthy individuals, and how the results of this test correspond to existing measures.

However, there is the assessment tool that speech/language therapists and neuropsychologists frequently use in acute clinical settings: Vasserman's scale (Vasserman, Dorofeeva, & Meerson, 1997). This scale is detailed, but it requires more time to administer (it can take more than 30 minutes) and it cannot be used by non-speech/language

specialists. Therefore, it should not be used as screening tools during the acute phase of stroke.

All such instruments have their advantages and disadvantages. In a situation when speech therapists are often not available in the hospital, short screening tests can be insufficient for understanding a person's deficit and this may lead to misdiagnosis. Meanwhile, neuropsychological batteries cannot be applied by non-speech/language professionals. There is a lack of simple tools for the separate assessment of language production and language comprehension. Also, the majority of existing tools do not report their psychometric parameters, which are essential for aphasia tests (Ivanova & Hallowell, 2013) or speech/language screening tests (Rohde et al., 2018). An instrument meeting all standards should allow neurologists to confirm their findings and to direct ways for the patient's future diagnostics and treatment.

The Aphasia Bedside Check (ABC; Lamers, Ruiters, El Hachoui, & Visch-Brink, 2020; Visch-Brink & El Hachoui, 2013) is successfully used in clinical practice in the Netherlands. Administration of this test does not require specific knowledge of speech/language disorders; however, it allows a specialist to assess the preservation of the patient's speech/language. Moreover, this test is developed as a small booklet that easily fits in a doctor's or other clinician's pocket, so it does not require any additional equipment. The ABC test is highly useful for diagnostics in the acute stage of stroke, especially in clinics where no professional speech/language therapists are employed. It allows any member of the medical staff to assess speech/language comprehension and production separately, enabling the identification of a patient's linguistic deficit(s).

Taking these facts into consideration, we developed a Russian version of the ABC (ABC-Ru). Compared to other instruments for aphasia screening in Russian, this test is short, linguistically motivated, and convenient for usage. Previous tests in Russian lack some of those qualities.

The items of the test will be discussed in detail in the further sections.

We conducted two studies to answer the following questions:

1. Does the ABC-Ru meet modern psychometric standards?
2. Is this test suitable for usage in the acute post-stroke period?

ABC-Ru: Description

The Aphasia Bedside Check consists of three parts. Part A is devoted to language comprehension, part B assesses language production, and part C includes complex additional tasks. Parts A and B form a core part of the test, while part C can be used optionally for a deeper understanding of the patient's deficit. Every part includes seven items.

Part A: Language Comprehension

This part consists of seven items. All items aim for fast evaluation of the patient's ability to comprehend language: 2 picture pointing tasks, 2 sentence judgment tasks, and 3 yes/no questions. These tasks are appropriate for the screening of auditory comprehension deficits. We used the Russian translations of the stimuli from the original Dutch version of the ABC with the preservation of its original aim and principle. Dutch and Russian native speakers who are professional linguists prepared the translation and adaptation of the stimuli.

A1. Sentence-picture matching. This task assesses morphosyntactic comprehension at the sentence level. The patient looks at two pictures. One of them depicts a cow, and another one depicts a bicycle. Then the patient hears the following instructions: "Touch a bicycle, and then touch a cow." The two words that are used for this task are comparable on relevant psycholinguistic variables (Akinina et al., 2016), as illustrated in Table 1.

A2. Spoken-word-to-picture matching. This task screens for the patient's ability to understand the meaning of a spoken word. The patient sees two pictures. One picture depicts an apple, and the other depicts a pie. While looking at the pictures, the patient hears the following instruction: "Touch the apple." The 'pie' is a semantic distracter for the target word 'apple.' Both words are phonologically simple, as shown in Table 2.

A3–A4. Sentence comprehension. This task tests the patient's sensitivity to the plausibility of auditorily and visually presented sentences. The examiner presents a picture and a sentence written above the picture. The examiner also reads the sentence aloud. The patient should decide whether the sentence fits the picture. The patient should say "yes" in case the sentence is correct and corresponds to the picture. If the sentence does not fit the picture, the patient should say "no."

In item A3, the patient looks at a picture showing a man going down the stairs and hears a sentence which does not fit the picture; see example 1. In this example, the case of the object is INSTR, while in the correct sentence, the case should be GEN (see example 2). The examiner asks if this sentence fits the picture. The expected answer to that question is "No."

(1) *Muzhchina spuskaetsya s lestnitsei
Man go down with stairs-INSTR
'The man is going down with the stairs'

(2) Muzhchina spuskaetsya s lestnitsy
Man go down from stairs-GEN
'The man is going down from the stairs'

In item A4, the patient sees a picture with a woman who is going to wash a car. The patient reads and hears the sentence with a compound predicate *Zhenshchina sobiraetsya pomyt' mashinu* 'The woman is going to wash the car.' The examiner asks if this sentence fits the picture. The expected answer to the examiner's question is "Yes."

A5–A7. Yes/No questions. This task is focused on the comprehension of sentences, tapping into different aspects of semantics. In this task, the examiner asks a question, and a patient should answer "yes" or "no." For details, see Table 3.

Part A: Scoring

Each item in part A is scored as 1 if the patient provides the expected answer. In all other cases, the response should be scored as 0. Slow answers, as well as cases of self-correction, should also be scored as 0.

Part B: Production

The production part of the ABC includes seven items to assess the patient's ability for language production,

Table 1. Stimuli Parameters for the Sentence-Picture Matching Task

Noun	Object Familiarity – Mean (SD)	Age of Acquisition – Mean (SD)	Frequency (ipm)	Length in Syllables	Length in Phonemes
Velosiped 'bike'	4.52 (.90)	1.66 (.65)	23.10	4	9
Korova 'cow'	4.19 (1.07)	1.20 (.40)	44.80	3	6

Note: ipm — instances per million words.

Table 2. Stimuli Parameters for the Spoken-Word-to-Picture Matching Task

Noun	Object Familiarity – Mean (SD)	Age of Acquisition – Mean (SD)	Frequency (ipm)	Length in Syllables	Length in Phonemes
Yabloko 'apple'	4.80 (.64)	1.07 (.26)	53.00	3	7
Pirog 'pie'	4.37 (.86)	1.47 (.67)	16.30	2	5

Note: ipm — instances per million words.

including object naming, word repetition, and reading words aloud. We preserved the principles of the original test and translated the tasks from the original Dutch version when possible. In some cases, we changed the stimuli to be more culture-specific and linguistically relevant for Russian.

B1–B3. Object naming. The patient is presented with two object pictures and one animal picture (one at the time) and their task is to name what they see in the picture. The complexity of the nouns increases item by item (see Table 4). We replaced the pictorial stimuli in item B1 (piano) to avoid ambiguity in the interpretation (two notions can refer to the object piano: *piano* and *fortepiano*). Instead, we used a picture of a guitar. We also replaced the picture for item B2 (*broccoli*) because its frequency is very low in Russian (0.7 instances per million words). In the ABC-Ru, we use the word *cherepakha*: ‘turtle’.

B4–B5. Repetition. In this task, the patient should repeat two abstract and phonologically complex nouns (see Table 5). We followed the principle of the original Dutch version, and the word used in item B4 was directly translated from Dutch to Russian. For item B5, we used a word

which is semantically and phonologically close to the Dutch original (*sollicitatie* ‘application’).

B6–B7. Reading aloud. Items B6 and B7 assess the patient’s ability to read phonologically complex words aloud. The two words for these items are chosen following the principle of the original ABC. For details, see Table 6.

Part B: Scoring

All items included in part B should be scored 1 if the answer is correct. The use of semantic and phonological paraphasias, jargon speech, perseverations, the response “do not know,” and no response all lead to scoring as 0.

Part C. Additional tasks

This part of the ABC is optional and can be applied when the examiner wishes to know more about the patient’s speech/language status. Part C consists of seven items of high linguistic complexity.

C1. Picture description. In this task, the examiner presents a picture, and asks a question: “Please, describe what is wrong in this picture?” We identified five elements of the picture which non-brain-damaged (NBD) individuals pointed out as strange or wrong, i.e., the umbrella (up-

Table 3. Stimuli Parameters for Yes/No Questions

Item	Question	Expected Answer	Parameters
A5	U stola est’ okna? ‘Does a table have windows?’	No	Sentence comprehension focused on semantics of the noun.
A6	Mozhno li s’est’ makarony? ‘Is it possible to eat spaghetti?’	Yes	Sentence comprehension focused on relations between a verb and a noun.
A7	Babochka — eto ptitsa? ‘Is a butterfly a bird?’	No	Sentence comprehension focused on category semantics.

Table 4. Stimuli Parameters for the Object Naming Task

Noun	Object Familiarity — Mean (SD)	Age of Acquisition — Mean (SD)	Frequency (ipm)	Length in Syllables	Length in Phonemes	Number of Consonant Clusters
Gitara ‘guitar’	4.06 (1.20)	1.98 (.77)	22.20	3	6	0
Cherepakha ‘turtle’	3.76 (1.33)	1.63 (.73)	6.30	4	8	0
Vertolyot ‘helicopter’	3.55 (1.36)	2.01 (.69)	38.70	3	8	1

Note: ipm — instances per million words.

Table 5. Stimuli Parameters for the Repetition Task

Noun	Frequency (ipm)	Length in Syllables	Length in Phonemes	Number of Consonant Clusters	Number of Articulatory Switches
Organizatsiya ‘organization’	312.4	6	11	1	2
Elektrichestvo ‘electricity’	20.2	4	13	2	3

Note: ipm — instances per million words.

side down), a duck, the pants, the cloud, and the boots. Patients score 1 point when they mention the umbrella and at least two out of the four other elements. If the patient names fewer than three elements or does not mention an umbrella, they get 0 points. This task assesses narrative speech production.

C2. Yes/No questions. This task contains three yes/no questions (see Table 7).

C3. Repetition. In this task, the patient should repeat three phonologically complex phrases (see Table 8).

C4. Reading aloud. In this task, the patient reads two phonologically complex words and one phrase aloud (see Table 9).

C5. Completion of a series of words. This task assesses a patient's word retrieval ability. The patient is presented with three series of words, each series of a different semantic category. Each series consists of three representatives of one category: a) *romashka* 'chamomile', *tyul'pan* 'tulip', *liliya* 'lily'; b) *schuka* 'pike', *sel'd* 'herring', *losos* 'salmon'; c) *kapusta* 'cabbage', *ogurets* 'cucumber', *morkov* 'carrot'. The examiner reads each series aloud and asks

the patient to complete it. If the patient names one representative of the category or the name of the category itself, they receive a score of 1. In all other cases, the patient gets 0 points.

C6. Verbal fluency task. This task tests the ability to retrieve words of a specific semantic category. There are three categories: animals, professions, cities. The examiner sets a timer and invites the patient to name as many representatives of a category as possible in one minute. The number of tested categories is up to an examiner. We calculated the threshold for each category as the number of items of a class (correctly) mentioned by non-brain-damaged speakers minus one standard deviation. The score is 1 if a patient names 15 representatives or more in the category of animals; 12 or more in the category of professions; 15 or more in the category of cities. If the patient names fewer items than the threshold, they get 0 points.

C7. Repetition of syllables series. This task allows us to estimate the presence or absence of articulation difficulties. In this task, the examiner asks a patient to repeat three series of syllables with explosive consonants differing

Table 6. Stimuli Parameters for Reading Aloud

Noun	Frequency (ipm)	Length in Syllables	Length in Phonemes	Number of Consonant Clusters	Number of Articulatory Switches
Administratsiya 'administration'	98.3	6	13	2	2
Pozdravlenie 'congratulations'	10.8	5	12	1	3

Note: ipm — instances per million words.

Table 7. Stimuli Parameters for Yes/No Questions (Part C)

Item	Question	Expected Answer	Parameters
C2a	Mozhno li dut' v baraban? 'Can you blow in a drum?'	No	Sentence comprehension focused on relations between a verb and a noun.
C2b	Kitay — eto strana? 'Is China a country?'	Yes	Sentence comprehension focused on semantics (i.e. world knowledge).
C2c	U slona est' klyuv? 'Does an elephant have a beak?'	No	Sentence comprehension focused on semantics of the noun.

Table 8. Stimuli Parameters for the Repetition Task (Part C)

Item	Noun	Length in Syllables	Length in Phonemes	Number of Consonant Clusters	Number of Articulatory Transitions
C3a	Futlyar dlya drel'i 'A drill case'	5	14	3	0
C3b	Dorozhnoe zagrazhdenie 'roadblock'	9	19	3	3
C3c	Opublikovat' oproverzhenie 'to publish a rebuttal'	11	24	3	4

in the place of articulation: *papapapapa* (with a labial consonant), *tatatatata* (with an alveolar consonant), and *kakakakaka* (with a velar consonant). The fourth series is made to elicit switches between those consonants: *pataka pataka pataka*. Every item is scored as 1 in the case of correct repetition, and 0 in all other cases. The maximum score for this task is 4.

Study 1

Study 1 aimed to standardize the ABC-Ru and to determine to what extent this test meets the required psychometric standards. We standardized the core test, including parts A (language comprehension) and B (language production). Part C is not necessary for aphasia screening but contains a set of additional tasks that a professional may or may not use for a deeper understanding of the patient's deficit. Therefore, this part has not been included in Study 1.

Method

Participants. Two groups of participants took part in the study. The clinical group consisted of people in the chronic post-stroke period: $N=80$, 30 females; mean age=60 years ($SD=11.5$, range 26–79); mean years of education=14 ($SD=2.5$, range 8–19); mean months post-onset=56 ($SD=66$, range 5–307). At admission to the Center for Speech Pathology and Neurorehabilitation (Moscow, Russia), these participants were assessed with standard tests by a professional speech language therapist (SLT). According to the SLT reports, 60 participants suffered from a variety of aphasia types (39 fluent, 21 non-fluent), and 20 were diagnosed with pure dysarthria without aphasia. None of the participants had a neurodegenerative disease or any psychiatric or neurological history.

The control group consisted of people who volunteered for the study: $N=120$, 61 females; mean age=43 years ($SD=19$, range 17–86); mean years of education=15 ($SD=2.6$, range 10–21). Participants of the control group were selected to match the people from the clinical group by age, gender, and education level.

All participants, except for one left-handed and one ambidextrous member of the clinical group and three left-handed members of the control group, were (premorbidly) right-handed. Participants were monolingual native speakers of Russian, had (corrected to) normal vision and hearing, and did not have any traumatic brain injuries in the last three years. None of them used medication which could affect cognitive functions. Detailed participant demographic information is presented in [Supplementary Online Material](#)

Table 1 (clinical group) and Table 2 (control group). All participants signed the informed consent form approved by the HSE Committee on Interuniversity Surveys and Ethical Assessment of Empirical Research (National Research University Higher School of Economics, Moscow, Russia).

Materials and procedure. All participants performed the ABC-Ru and the Token Test (Token Test App: Bastiaanse, Akinina, Satoer, & de Kok, 2020; Russian version: Akinina, Buivolova, Soloukhina, & Bastiaanse, 2019) in a randomized order. The Token Test is a test for detecting language disorders (De Renzi & Faglioni, 1978). For the Token Test, participants see squares and circles of different colors, shapes, and sizes and hear instructions on how they should manipulate the figures (for example, *Touch the black square; Touch the yellow circle and the green square; Touch all the circles, except the green one*). There are five series of commands which become more complex series by series. Participants were tested in a quiet room by a clinical linguist. All spoken answers to the ABC-Ru were recorded after acquiring additional oral consent, and responses were scored on a paper form. All answers on the Token Test were automatically scored in the App.

For evaluating the ABC-Ru as a *reliable* test (Greenhalgh, 1997), we measured its sensitivity, specificity, positive and negative predictive values, as well as positive and negative likelihood ratios. The sensitivity of the test allows diagnosing people with aphasia as such. The specificity of the test shows how well the test allows one to exclude people without aphasia. Positive/negative predictive values estimate how probable it is that when the results of the test are positive/negative, people have or do not have aphasia. The positive/negative likelihood ratio reflects the probability of finding positive/negative results in a person with/without aphasia. We estimated all these parameters according to the formula proposed by Greenhalgh (1997).

We measured the *concurrent validity* (Ivanova & Hallowell, 2013) of the test by using the correlation between the results obtained by the ABC-Ru with the results of the valid and reliable Token Test (Akinina, Bastiaanse, & Soloukhina, 2019, Akinina, Buivolova, Soloukhina, & Bastiaanse, 2019). *Construct validity* was measured by a comparison of the results on the ABC-Ru and SLT assessments.

For measuring *inter-rater reliability*, four independent raters scored the recorded answers of 30 participants (with aphasia and/or dysarthria) chosen randomly from the clinical group.

To evaluate *internal consistency*, we measured how each item correlates with the final score using the Kuder-

Table 9. Stimuli Parameters for Reading Aloud (Part C)

Item	Noun	Length in Syllables	Length in Phonemes	Number of Consonant Clusters	Number of Articulatory Transitions
C4a	Abonement 'subscription'	4	9	1	3
C4b	Laboratoriya 'laboratory'	6	11	0	4
C4c	Neveroyatnaya konstruktsyya 'incredible construction'	10	22	3	6

Richardson formula 20. For item analysis, we determined item difficulty to ensure that the test includes items of different complexity. We measured the percentage of correct answers for each item in a representative sample of people with aphasia of different types and severity (60 people with aphasia from the clinical group) and we compared them to the results of healthy individuals. Moreover, we estimated item discriminability to evaluate the degree to which every item refers to the same construct with the overall score.

To measure the effect of *demographic characteristics* on test performance, we correlated the results of the ABC-Ru with age and educational background in both the clinical and control groups, and the results on the ABC-Ru with the time post-onset in the clinical group. Additionally, we compared the performance of participants with different types of aphasia (fluent vs. non-fluent).

Results

The mean ABC-Ru score in the group of people with aphasia was 10.50 ($SD=3.29$, range 2–14), and the mean Token Test score was 20.16 ($SD=9.02$, range 0.50–36). In the group of participants with dysarthria ($N=20$), the mean overall score on the ABC-Ru was 12.50 ($SD=2.08$, range 7–14), and the mean Token Test score was 29.05 ($SD=7.86$, range 7.50–36). In the control group, participants performed on the ABC-Ru with an overall score of 13.54 ($SD=0.74$, range 8–14) and on the Token Test with an overall score of 34.55 ($SD=1.90$, range 27–36).

We defined the cut-off score to allow for differentiation between performance of those with and without speech/language disorders as equal to the 5th percentile of the performance in the control group. For the core part of the test, including parts A and B, the cut-off score was 13 out of 14 points. The same cut-off score was determined for the original Dutch version of the test.

The results for sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios according to the formula discussed in Greenhalgh (1997) are given in Table 10. We used the participants' results on the ABC-Ru and on the Token Test as a golden standard measure. We included the results of the control group and the results of the group of people with SLT-diagnosed aphasia, because the Token Test is supposed to detect language disorders but not speech disorders such as dysarthria. The ABC-Ru is sensitive to articulation problems, so it was not suitable for a comparison. Hence, to avoid ambiguity, we did not include the data obtained in the group of dysarthric participants. The results are given in Table 10.

The results showed that the ABC-Ru can be considered as sensitive (72% of people with aphasia were diagnosed as such) and highly specific (96% of participants were estimated correctly as not having aphasia). We found that the test has high positive (.88) and negative (.89) predictive values, referring to the probability that people will be diagnosed correctly as having aphasia (positive) or not having aphasia (negative), respectively. The positive likelihood ratio (17.35) is due to a very high probability of find

Table 10. Reliability Measures of the ABC-Ru

	TT<29 (N=53)	TT≥29 (N=121)
ABC-Ru<13 (n=42)	A True positive (38)	B False positive (5)
ABC-Ru≥13 (n=132)	C False negative (15)	D True negative (116)

	Value	95% confidence intervals
Sensitivity A/(A+C)	.72	.58–.83
Specificity D/(B+D)	.96	.91–.99
Positive predictive value A/(A+B)	.88	.75–.96
Negative predictive value D/(C+D)	.89	.82–.93
Likelihood ratio of a positive test sensitivity / (1-specificity)	17.35	7.23–41.61
Likelihood ratio of a negative test (1-sensitivity) / specificity	.30	.19–.45

Note: ABC-Ru<13: the score on the ABC-Ru is lower than 13, indicating speech/language problems. ABC-Ru≥13: the score on the ABC-Ru is 13 or higher, indicating the absence of speech/language problems. TT<29: the Token Test score is lower than 29, indicating aphasia. TT≥29: the Token Test score is equal to or higher than 29, indicating the absence of aphasia.

a positive result in a person with aphasia, and the negative likelihood ratio of 0.29 refers to a moderate probability of finding a negative result on the ABC-Ru in a person without aphasia (McGee, 2002).

To identify the concurrent validity with the golden standard instrument, we measured the correlation between the ABC-Ru and the Token Test results in the three groups. In the control group, the Pearson's correlation coefficient was low but significant ($r=.356, p<.001$). In the group of participants with aphasia, Pearson's correlation was moderate and significant ($r=.592, p<.001$). In the group of people with dysarthria, the results on the ABC-Ru and the Token test did not correlate significantly ($r=-.109, p=.645$). These results can be explained by the fact that the Token Test is meant to diagnose language disorders (i.e., aphasia) and, hence, patients with dysarthria should perform relatively normally on this test. The ABC-Ru is meant to detect both speech and language disorders. This shows that results on both tests correlate well in a group of people with aphasia, indicating that the ABC-Ru is about as good for aphasia detection as the Token Test.

To identify the construct validity of the ABC-Ru, we calculated the correlation coefficient (Pearson) between

the performance on the ABC-Ru and the severity of speech/language disorders (range 1–6 from mild to very severe disorder) in the clinical group, measured by professional SLTs using standardized methods. We found a moderate and significant correlation between those two measures ($r=-.481, p<.001$), meaning that the ABC-Ru allows us to detect the presence of speech/language disorders and to estimate their severity. The coefficient is negative due to the opposite direction of the scales.

For measuring inter-rater reliability, four independent raters (two were trained neurolinguists, and the other two were trained students from the HSE linguistics department) evaluated the recordings of 30 participants randomly selected from the clinical group. Raters scored participant performance on the tasks from core parts A and B (except for items A1 and A2, where the participant points out in the pictures what makes an evaluation based only on audio recording impossible). After the exclusion of these two items from the inter-rater reliability analysis, the maximal overall score for inter-rater reliability was 12 with the maximal score of 5 for part A and 7 for part B. The mean scores were 9.90 ($SD=2.41$, range 3–12) for Rater 1; 9.78 ($SD=2.21$, range 3–12) for Rater 2; 9.77 ($SD=2.29$,

Table 11. Inter-Item Complexity

Item	NBD — Difficulty Mean (<i>SD</i>)	PWA — Difficulty Mean (<i>SD</i>)	Difficulty Interpretation	Discriminability	Discriminability Interpretation
A1	.99 (.09)	.93 (.25)	Easy	.15	Fair
A2	1.00 (.00)	.97 (.18)	Easy	.10	Fair
A3	.73 (.47)	.37 (.49)	Hard	.55	Good
A4	.95 (.24)	.92 (.28)	Easy	.10	Fair
A5	.98 (.13)	.82 (.39)	Moderate	.35	Good
A6	.99 (.09)	.87 (.34)	Easy	.25	Fair
A7	.99 (.09)	.93 (.25)	Easy	.20	Fair
B1	.99 (.09)	.78 (.42)	Moderate	.60	Good
B2	1.00 (.00)	.67 (.48)	Moderate	.50	Good
B3	.98 (.13)	.70 (.46)	Moderate	.75	Good
B4	1.00 (.00)	.67 (.48)	Moderate	.90	Good
B5	1.00 (.00)	.65 (.48)	Moderate	.90	Good
B6	1.00 (.00)	.67 (.48)	Moderate	.85	Good
B7	.99 (.09)	.57 (.50)	Moderate	.75	Good

Note: NBD — group of non-brain-damaged individuals; PWA — people with aphasia.

range 4–12) for Rater 3; and 10 ($SD=2.18$, range 3–12) for Rater 4. To assess the degree of consistency of scorings between the raters, we used a two-way random, absolute, average-measures intraclass correlation (ICC; McGraw & Wong, 1996). The found intraclass correlation of .955 (95% CI: .922–.977; $F(29, 89.4) = 22.1, p < .001$) can be considered as excellent (Koo & Li, 2016).

We measured the internal consistency of the test using the Kuder-Richardson formula (KR-20) because the data were scored dichotomously (Cicchetti, 1994; Ivanova & Hallowell, 2013). The results indicate that the ABC-Ru can be considered as homogenous (KR-20 = .850) (Pattock, 2004). The inter-item analysis showed that the items of the test are balanced in terms of difficulty: five items can be considered as easy, seven as moderately difficult, and one item as hard (Office of Educational Assessment, 2020). Discriminability of every single item varies from fair to good (Office of Educational Assessment, 2020). For item analysis details, see Table 11.

To identify the influence of demographic characteristics on the ABC-Ru results, we applied a Pearson's correlation test a) between participant age and ABC-Ru results in the control ($r = -.142, p = .120$) and clinical ($r = -.113, p = .320$) groups; b) between educational level and ABC-Ru results in the control ($r = .145, p = .105$) and clinical ($r = .106, p = .349$) groups; and, finally, c) between the time post-onset and ABC-Ru results in the clinical group ($r = -.171, p = .127$). We also divided the control group into two age subgroups (17–45 years old, $n = 64$; 45–86 years old, $n = 56$) and found no difference between their performance (Wilcoxon signed-rank test: $Z = -1.403, p = .161$). Two age groups of participants from the clinical group (26–59 years old, $n = 34$; 60–86 years old, $n = 46$) performed similarly as well ($Z = -.238, p = .778$). Finally, there was no performance difference between individuals with fluent ($n = 38$) and non-fluent ($n = 22$) aphasia ($Z = -1.036, p = .300$). As expected, there was a significant difference between the performance of the groups of people with aphasia and people with dysarthria ($Z = -3.296, p = .001$). In the group of people with dysarthria, there was no significant difference between performance on comprehension and production tasks ($Z = -.524, p = .600$).

Interim Discussion

In this first study, we showed that the ABC-Ru is a screening test for speech/language disorders which meets the required psychometric standards. This new test can be qualified as reliable with high sensitivity, specificity, positive and negative predictive values, and appropriate likelihood ratios. Moderate and significant correlations for concurrent validity in the group of participants with aphasia demonstrate that the ABC-Ru can help to detect aphasia as effectively as the Token Test. The construct validity of the ABC-Ru is moderate, implying that the ABC-Ru allows us to estimate speech/language disorder severity as effectively as the standard tests used by SLTs. Two independent raters scored the results of people with and without speech/language disorders with a high agreement. This means that the test results can be scored reliably by any rater.

Item analysis showed the homogeneity of the test, and that the ABC-Ru includes items of varying difficulty (i.e.,

hard, moderate, and easy). Testing in the control group showed that people without brain damage performed almost perfectly on all items (95–100% correct answers), except for one item (A3). However, their performance on that item (73% correct responses) was significantly above chance, and we decided to keep the item as the only highly difficult item. Moreover, item analysis showed that all items that were hard or of moderate difficulty have a good discriminability, meaning that they correspond better to the overall score than relatively easy items. Taking all these results into account, we propose that the final set of five 'easy', seven moderately difficult, and one 'hard' items provides an appropriate range to make a distinction within participants with disorders of different severity (Ivanova & Hallowell, 2013).

Finally, we found that demographic characteristics do not influence ABC-Ru scores. Moreover, there was no difference in performance in the groups of people with different types of aphasia, which means that the test can help to detect various kinds of aphasia with the same reliability. A significant difference was found between the performance in the group of people with aphasia versus pure dysarthria. The overall ABC-Ru score of the latter group was higher than in the group of people with aphasia, though below the threshold, and there was no significant difference in their performance on production tasks versus comprehension tasks. The results of participants with pure dysarthria on the Token Test were above threshold. We cannot claim that the ABC-Ru allows us to differentiate between aphasia and articulation disorders. However, the combination of the ABC-Ru and Token Test can give further direction for diagnostics.

Study 2

After standardization of the test in the population of people with stable speech/language status, we investigated whether the test is suitable for usage in the acute stage in clinical settings. We estimated how well the results obtained on the ABC-Ru correspond to speech/language disorder severity assessed by SLTs and to the score on another standardized screening test, the Russian Aphasia Rapid Test (Russian ART; Buivolova et al., 2020).

Method

Participants. A clinical group of 20 individuals (7 females; mean age = 61.2 years, $SD = 13.9$, range 32–85) took part in the study. All participants suffered from a stroke and were in the acute post-stroke period (mean time post-onset 10 days, $SD = 4.8$, range 2–23 days), and they did not have neurodegenerative disorders. They were admitted to the neurology department of the Moscow City Clinical Hospital No31. According to the SLT assessment, 10 participants had aphasia, 6 participants had pure dysarthria, and 4 participants did not have any speech and/or language disorders. Participant demographic information is presented in [Supplementary Online Material Table 3](#).

Materials and procedure. All participants were tested by a clinical linguist or professional SLT while in bed, using the Russian *Aphasia Rapid Test* (ART; Buivolova et al., 2020) and the core part (parts A and B) of the ABC-

Ru. The order in which the two tests were administered was randomized. The Russian ART was chosen because it is a valid and standardized screening test for speech/language disorders and can be considered as a golden standard screening test for Russian. It consists of six tasks, assessing the understanding of instructions, repetition of three words and one sentence, naming of three objects, scoring of dysarthria, and a semantic fluency task. On the same day, the SLT provided a detailed speech/language assessment with the Vasserman's scale, a test for neuropsychological evaluation (Vasserman et al., 1997). According to the results, the SLT rated the speech/language disorder severity on a scale from 1 to 6. All participants or their relatives signed the informed consent for participation in the project and allowed personal data processing at admission to the hospital. A professional started to test patients after an explanation of the procedure and obtaining oral consent for participation in the study.

Results

The participants performed the Russian ART with a mean score of 7.60 out of 26 points ($SD=7.50$, range 0–23) and the ABC-Ru with a mean score of 10.25 out of 14 ($SD=3.58$, range 2–14). The correlation coefficient (Spearman) of the results of the two tests was high and significant ($\rho=-.911$, $p<.001$). Moreover, the results on the ABC-Ru highly correlated with the severity of the speech/language disorder according to the SLT assessment ($\rho=-.835$, $p<.001$). The coefficients are negative because the tests have oppositely directed scales. The correlation between the Russian ART and SLT assessment was also high and significant ($\rho=.856$, $p<.001$).

People with aphasia and/or dysarthria performed the ABC-Ru with a mean score of 9.50 ($SD=3.63$, range 2–14), while people who suffered from stroke without any accompanying speech/language disorders performed the ABC-Ru above the threshold with a mean score of 13.25 ($SD=0.50$, range 13–14). The difference in performance in those two groups is significant according to the results of the Wilcoxon signed-rank test ($Z=-3.308$, $p<.001$).

Interim Discussion

The second study shows that the ABC-Ru is as suitable for speech/language screening as the Russian ART. They measure the same construct, and both correlate well with the results of the SLT assessment. Although both tests can be successfully used for speech/language pathology screening, the ABC-Ru allows for assessment of a wider range of patient abilities (e.g., sentence comprehension and reading aloud) during a comparable amount of time. Moreover, the ABC-Ru allows us to assess language comprehension and production separately and to use the additional tasks for more detailed screening if necessary. The results show that the ABC-Ru allows us to differentiate between post-stroke patients with and without speech/language disorders.

General Discussion

We conducted two studies to test whether the Russian version of the Aphasia Bedside Check meets modern psychometric standards and can be used for the screening

of speech/language pathologies in the acute post-stroke period. The ABC-Ru allows us to screen for problems at the levels of language comprehension and language production, and to perform additional tasks that can be informative for the development of a further rehabilitation course for each patient. Administration of the test does not require any specific skills and can be performed not only by speech/language pathologists or neuropsychologists but also by neurologists and other members of the neurological ward staff. Our results show that the test meets standards and can be used in the acute post-stroke period.

Our study has some limitations. Firstly, due to the COVID-19 pandemic, we were unable to collect data for a test-retest analysis. This would have allowed us to understand whether the results of the test are stable over time. However, this is not crucial for the ABC-Ru, because this test is supposed to be used for screening in the acute post-stroke period and not to measure improvement. It is known that spontaneous language recovery during the acute stage of stroke is possible (Holland & Fridriksson, 2001), so it cannot be proven whether changes in test results are caused by low test-retest reliability or by spontaneous recovery processes. To estimate whether the test-retest reliability is high, it will be necessary to compare the results in a group of people with chronic aphasia and people with acute aphasia, and to define whether there are significant differences between the performance at two time points in those two groups. This should be done in a future study.

The second limitation is devoted to the small number of participants in Study 2. However, in Study 1, the ABC-Ru was administered with 120 healthy individuals, 80 people with chronic aphasia, and 20 people with chronic dysarthria. In Study 2, the results obtained in the group of 20 people in the acute post-stroke period confirmed findings from the chronic population and showed a high correlation with existing measures. Thus, for the current study, the sample size can be considered as sufficient.

Lastly, the main purpose of the current study was to standardize the core part of the ABC-Ru, including parts A (language comprehension) and B (language production). The optional part C with the additional items has been adapted to Russian, but since it is an optional part, it was not the focus of the present study.

As for future perspectives, the ABC-Ru shows promising potential for being used in the neurosurgical population. As of now, there is no short and valid test for measuring the existence of aphasia and the severity of pre- and post-operative speech/language disorders, especially in patients with severe speech/language problems. That is why the ABC-Ru should be standardized and validated not only in the post-stroke population but in people with brain tumors and traumatic brain injuries.

To sum up, the standardization of the Aphasia Bedside Check for Russian showed promising results. This test can be considered as a reliable instrument for speech/language screening of patients in the acute post-stroke period.

References

- Akhutina, T. (2015). Luria's classification of aphasias and its theoretical basis. *Aphasiology*, 30(8), 878–897. <https://doi.org/10.1080/02687038.2015.1070950>

- Akinina, Y., Bastiaanse, R., & Soloukhina, O. (2019). *Automated language deficit diagnostics with The Token Test App: Test-retest reliability and practice effects*. Paper presented on Conference on Serious Games and Applications for Health, Kyoto, Japan, 2019, August 5–7.
- Akinina, Y., Buivolova, O., Soloukhina, O., & Bastiaanse, R. (2019). Psychometric properties of the Token Test App. *Stem-, Spraak- en Taalpathologie*, 24(Supplement), 74–76. Retrieved from <https://sstp.nl/article/view/35738/33292>.
- Akinina, Y., Grabovskaya, M., Vechkaeva, A., Ignatyev, G., Isaev, D., & Khanova, A. (2016). Библиотека психолinguистических стимулов: новые данные для русского и татарского языка [Psycholinguistic stimulus database: New data for Russian and Tatar]. In Y. Alexandrov, & K. Anokhin (Eds.), *Seventh International Conference on Cognitive Science: Abstracts. Svetlogorsk, June 20-24 2016* (pp. 93–95). Moscow: Institut Psikhologii RAN. (In Russian).
- Azuar, C., Leger, A., Arbizu, C., Henry-Amar, F., Chomel-Guil-laume, S., & Samson, Y. (2013). The Aphasia Rapid Test: An NIHSS-like aphasia test. *Journal of Neurology*, 260(8), 2110–2117. <https://doi.org/10.1007/s00415-013-6943-x>
- Bastiaanse, R., Akinina, Y., Satoer, D., & de Kok, D. (2020). *The Token Test App, Second Edition*. Groningen, Netherlands: Retrieved from <https://play.google.com/store/apps/details?id=nl.rug.getc.tokenest>.
- Buivolova, O., Vinter, O., Bastiaanse, R., & Dragoy, O. (2020). The Aphasia Rapid Test: Adaptation and standardisation for Russian. *Aphasiology*, 1–15. <https://doi.org/10.1080/02687038.2020.1727836>
- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex*, 14(1), 41–49. [https://doi.org/10.1016/s0010-9452\(78\)80006-9](https://doi.org/10.1016/s0010-9452(78)80006-9)
- El Hachoui, H., Visch-Brink, E. G., de Lau, L. M. L., van de Sandt-Koenderman, M. W. M. E., Nouwens, F., Koudstaal, P. J., & Dippel, D. W. J. (2016). Screening tests for aphasia in patients with stroke: A systematic review. *Journal of Neurology*, 264(2), 211–220. <https://doi.org/10.1007/s00415-016-8170-8>
- Greenhalgh, T. (1997). How to read a paper: Papers that report diagnostic or screening tests. *British Medical Journal*, 315(7107), 540–543. <https://doi.org/10.1136/bmj.315.7107.540>
- Holland, A., & Fridriksson, J. (2001). Aphasia management during the early phases of recovery following stroke. *American Journal of Speech-Language Pathology*, 10(1), 19–28. [https://doi.org/10.1044/1058-0360\(2001\)004](https://doi.org/10.1044/1058-0360(2001)004)
- Institute for Health Metrics and Evaluation (IHME) (2018). *Russia profile*. Seattle, WA: IHME, University of Washington. Retrieved from <http://www.healthdata.org/Russia> (09.06.2020).
- Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*, 27(8), 891–920. <https://doi.org/10.1080/02687038.2013.805728>
- Koo, T. K., & Li, M. Y. (2016). Cracking the code: Providing insight into the fundamentals of research and evidence-based practice. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lamers, E., Ruiter, M., El Hachoui, H., & Visch-Brink, E. G. (2020). Diagnostiek in de acute fase van een beroerte. Een eerste normering van de Afasie Bedside Check (ABC). *Nederlands Tijdschrift voor Logopedie*, 92, 20–25.
- Lucot, C., Koleck, M., Laurent, K., Darrigrand, B., Bordes, J., Joseph, P.-A., Dehail, P., & Mazaux, J.-M. (2013). Quality of life in stroke patients with aphasia. *Annals of Physical and Rehabilitation Medicine*, 56, e46. <https://doi.org/10.1016/j.rehab.2013.07.018>
- McGee, S. (2002). Simplifying likelihood ratios. *Journal of General Internal Medicine*, 17(8), 647–650. <https://doi.org/10.1046/j.1525-1497.2002.10750.x>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989x.1.1.30>
- National Institute of Neurological Disorders and Stroke (U.S.). (2011). *NIH stroke scale*. Bethesda, Md.: National Institute of Neurological Disorders and Stroke, Dept. of Health and Human Services, USA. Retrieved from https://www.ninds.nih.gov/sites/default/files/NIH_Stroke_Scale_Booklet.pdf.
- National Stroke Foundation (2010). *Clinical guidelines for stroke management 2010*. Melbourne, Australia.
- Office of Educational Assessment (2020). *Understanding item analyses*. Seattle, WA: University of Washington. Retrieved from <https://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis>.
- Patock, J. (2004). Exam scores: How to interpret your statistical analysis reports. Arizona State University University Testing and Scanning Services. Retrieved from https://uoeee.asu.edu/sites/default/files/docs/Guide_stat_analy_exam_scores.pdf.
- Rohde, A., Worrall, L., Godecke, E., O'Halloran, R., Farrell, A., & Massey, M. (2018). Diagnosis of aphasia in stroke populations: A systematic review of language tests. *PLOS ONE*, 13(3), e0194143:1–17. <https://doi.org/10.1371/journal.pone.0194143>
- Scherbakova, M. M., & Kotov, S. V. (2015). Screening method for examination of patient with aphasia. *Consilium Medicum*, 17(2), 8–11. (In Russian). <https://doi.org/10.26442/2075-1753.2015.2.8-11>
- Vasserman, L. I., Dorofeeva, S. A., & Meerson, Y. A. (1997). *Metody neiropsichologicheskoy diagnostiki: Prakticheskoe rukovodstvo [Neuropsychological assessment methods: Practical guide]*. Saint Petersburg: Stroylespechat. (In Russian).
- Visch-Brink, E., & El Hachoui, H. (2013). *Afasie Bedside Check*. Rotterdam, Netherlands: Erasmus MC.

МЕТОДЫ

Адаптация скринингового теста «Афазия: быстрая верификация» для русского языка

Ольга Витальевна Буйволова

Центр языка и мозга, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия; Научно-исследовательский центр медицинской реабилитации, ФГБУ «Федеральный центр мозга и нейротехнологий» ФМБА, Москва, Россия

Ивонна Рулинтье**Мария Бастиаансе**

Центр языка и мозга, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия; Центр языка и познания Гронингена, Университет Гронингена, Гронинген, Нидерланды

Ольга Викторовна Драгой

Центр языка и мозга, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия; Научно-исследовательский центр медицинской реабилитации, ФГБУ «Федеральный центр мозга и нейротехнологий», Москва, Россия

Оксана Борисовна Винтер

Отделение неврологии, Городская клиническая больница №31, Москва, Россия

Виктория Андреевна Позднякова

Центр языка и мозга, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия

Анастасия Александровна Самоукина

Центр языка и мозга, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия

Анастасия Андреевна Шляхова

Центр языка и мозга, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия

Эви Виш-Бринк

Департамент нейрохирургии, Медицинский центр Эразма, Роттердам, Нидерланды

Аннотация. «Афазия: быстрая верификация» (АБВ) — это адаптированный для русского языка скрининговый тест, который позволяет медицинскому персоналу неврологического отделения оценить наличие речевых нарушений у пациентов в первые дни после инсульта. Мы провели два исследования, направленные на определение соответствия теста АБВ современным психометрическим стандартам. В исследовании 1 докладываются результаты стандартизации теста в группе людей с хроническими речевыми нарушениями ($N=80$) и в когорте неврологически здоровых носителей русского языка ($N=120$). В исследовании 2 представлены результаты валидизации в группе людей в остром постинсультном периоде как с наличием речевых нарушений, так и без них. Результаты исследования показали, что скрининговый тест АБВ является валидным инструментом, который может быть использован для скрининга речевых нарушений в остром периоде инсульта.

Контактная информация: Ольга Витальевна Буйволова, obuivolova@hse.ru, 105066, Москва, Старая Басманная, 21/4, каб. 510; Ивонна Рулинтье Мария Бастиаансе, ybastiaanse@hse.ru; Ольга Викторовна Драгой, odragoy@hse.ru; Оксана Борисовна Винтер, oksvinter@mail.ru; Виктория Андреевна Позднякова, vpozdneyakova@hse.ru; Анастасия Александровна Самоукина, nastya.samoukina@gmail.com; Анастасия Андреевна Шляхова, aseshly32@gmail.com; Эви Виш-Бринк, e.visch-brink@erasmusmc.nl.

Ключевые слова: скрининговый тест, стандартизированная методика, острый период инсульта, афазия, русский язык

© 2020 Ольга Витальевна Буйволова, Ивонна Рулинтье Мария Бастиансе, Ольга Викторовна Драгой, Оксана Борисовна Винтер, Виктория Андреевна Позднякова, Анастасия Александровна Самоукина, Анастасия Андреевна Шляхова, Эви Виш-Бринк. Данная статья доступна по лицензии [Creative Commons "Attribution" \(«Атрибуция»\) 4.0. всемирная](https://creativecommons.org/licenses/by/4.0/), согласно которой возможно неограниченное распространение и воспроизведение этой статьи на любых носителях при условии указания авторов и ссылки на исходную публикацию статьи в данном журнале в соответствии с канонами научного цитирования.

Благодарности. Мы благодарим Р.А. Черемина и Г.Е. Иванову за помощь в организации исследования и возможность работать с пациентами Центра патологии речи и нейрореабилитации и Городской клинической больницы № 31. Мы признательны нашим коллегам из Центра языка и мозга (ВШЭ) за ценные комментарии при подготовке статьи. Мы хотим также выразить благодарность всем участникам этого исследования. Работа выполнена на базе Центра языка и мозга НИУ ВШЭ при финансовой поддержке гранта Правительства Российской Федерации № 14.641.31.0004.

Статья поступила в редакцию 31 июля 2020 г. Принята в печать 25 сентября 2020 г.

Введение

Инсульт является второй ведущей причиной смерти в России и восьмой по распространенности причиной получения инвалидности (Institute for Health Metrics and Evaluation, 2018). Одним из возможных последствий инсульта наряду с двигательными нарушениями является афазия. Афазия — нарушение, ведущее к проблемам с порождением и пониманием речи, что негативно сказывается на качестве жизни человека (Lucot et al., 2013). Для того, чтобы спрогнозировать исход заболевания и минимизировать его последствия, необходимо диагностировать речевые нарушения как можно раньше.

Для оценки речевой функции пациентов, поступающих в неврологические отделения, используются скрининговые тесты (National Stroke Foundation, 2010). Скрининг является важным этапом для дальнейшей диагностики и составления программы реабилитации. В остром периоде заболевания речь обычно оценивается с использованием общих скрининговых шкал, таких как National Institutes of Health Stroke Scale (NIHSS, National Institute of Neurological Disorders and Stroke [U.S.], 2011), а также с помощью инструментов, предназначенных специально для скрининга речевых нарушений. Такие инструменты широко используются в мировой клинической практике (El Hachoui et al, 2017; Rohde et al., 2018).

В российской клинической практике шкала NIHSS рекомендована для скрининга состояния пациента в первые дни от начала заболевания. Эта шкала позволяет оценить состояние различных функций, включая речь. Однако для детального речевого скрининга данная шкала не предназначена. Быстрый тест на афазию (БТА, Azuar et al., 2013; Русская версия: Buivolova et al., 2020) — стандартизированный лингвистический скрининговый тест, разработанный на основе NIHSS. Обе шкалы (БТА и NIHSS) просты как для исследователя, так и для пациента и подходят для начального скрининга состояния речевой функции. Однако детальной информации о ядре речевого дефицита с помощью этого теста получить невозможно.

Еще один скрининговый тест, который используется в российских клиниках, — «Методика скрининг-диагностики больных с афазией» (Щербакова, Котов, 2015). Этот тест был разработан для диагностики речевых нарушений. Он позволяет неврологам, логопедам и нейропсихологам определить тип афазии в соответствии с классификацией А.Р. Лурии (Akhtutina, 2016). Выполнение теста занимает небольшое количество времени (от 5 до 30 минут). Этот инструмент может быть использован для скрининга и отвечает запросам специалистов. Однако авторы теста не приводят информации о стандартизации теста, был ли он нормирован в группе неврологически здоровых испытуемых, а также как соотносятся результаты, полученные при выполнении этого теста, с результатами выполнения других существующих инструментов, используемых для скрининга.

Логопедами и нейропсихологами в неврологических отделениях для больных с острым нарушением мозгового кровообращения (ОНМК) широко используется шкала Л.И. Вассермана (Вассерман и др., 1997). К преимуществам этого теста относится его детализированность, к недостаткам — продолжительность (он занимает более 30 минут) и невозможность проведения специалистами без логопедической подготовки. Поэтому использование этой шкалы в остром периоде может быть затруднительным.

Все вышеперечисленные инструменты имеют преимущества и недостатки. В условиях нехватки логопедов в больницах короткие скрининговые тесты могут быть недостаточны для определения речевого дефицита, что может привести к неправильному диагнозу. При этом батареи нейропсихологических тестов не могут применяться неспециалистами в этой области. На данном этапе не хватает простых инструментов для оценки восприятия и порождения речи. Кроме того, авторы существующих тестов не всегда достаточно подробно описывают результаты стандартизации тестов, хотя это является важной частью разработки тестов для определения афазии (Ivanova, Hallowell, 2013) и речевых скрининговых тестов в целом (Rohde et al., 2018). Таким образом, необходимой

представляется разработка теста для русского языка, который соответствовал бы современным психометрическим стандартам и позволял бы неврологам определять направление дальнейшей диагностики и реабилитации и находить подтверждение поставленному диагнозу.

Тест Aphasia Bedside Check (ABC; Lamers et al. 2020; Visch-Brink, El Hachoui, 2013) широко используется в клинической практике в Нидерландах. Проведение теста не требует специальных знаний в области речевых нарушений, но позволяет медицинскому работнику оценить сохранность речевой функции пациента. Кроме того, этот тест разработан в виде буклета карманного размера и не предполагает использования дополнительного оборудования. Все это делает его крайне полезным для диагностики речевых нарушений у пациентов в остром периоде инсульта, особенно в больницах, где нет профессиональных логопедов. ABC позволяет любому члену медицинского персонала оценить порождение и восприятие речи, что может углубить понимание речевого дефицита пациента.

Принимая во внимание эти преимущества, мы разработали русскоязычную версию теста ABC, которая на русском языке получила название «Афазия: быстрая верификация» (АБВ). По сравнению с другими инструментами для определения наличия афазии у носителей русского языка, этот тест непродолжителен, лингвистически мотивирован и прост в использовании. Подробно стимульный материал теста обсуждается в следующих разделах.

Мы провели два исследования, отвечающие на следующие вопросы:

1. Соответствует ли тест АБВ современным психометрическим стандартам?
2. Применим ли тест для работы с пациентами в острой стадии инсульта?

АБВ: описание

Тест «Афазия: быстрая верификация» состоит из трех частей. Часть А оценивает понимание речи, часть Б — порождение, в часть В включены задания повышенной сложности. Части А и Б составляют ядро теста, часть В является дополнительной и может проводиться для углубления понимания речевого дефицита пациента. Каждая часть содержит семь заданий.

Часть А: понимание речи

Часть А представлена семью заданиями, каждое из которых направлено на быструю оценку способности пациента к пониманию речи. В двух заданиях пациенту необходимо выполнить инструкции или соотнести услышанное слово с рисунком. В других двух заданиях — оценить грамматическую правильность предложения. И, наконец, в последних трех заданиях — дать ответ «да» или «нет» на вопросы, связанные с восприятием смысла предложений. Эти задания подходят для диагностики нарушений восприятия речи на слух. В русской версии мы сохранили общий принцип построения и задачи оригинальной версии теста, по возможности сделав перевод с нидерландского языка, а в тех заданиях, где это было невозможно, мы подобрали релевантные аналоги в русском языке. Перевод и адаптация заданий осуществлялись лингвистами, носителями нидерландского и русского языков.

А1. Соотнесение предложения и изображений. Задание оценивает восприятие морфосинтаксиса на уровне предложения. Пациенту предъявляется две фотографии. На одной из них изображена корова, на другой — велосипед. После этого звучит инструкция: «Дотроньтесь сначала до велосипеда, а затем — до коровы». Два стимульных слова, использованных в задании, сопоставимы по психолингвистическим параметрам (Akinina et al., 2016), что показано в таблице 1.

Таблица 1. Параметры стимулов для задания на сопоставление предложения и картинки

Существительное	Знакомство с объектом: среднее (<i>SD</i>)	Возраст усвоения: среднее (<i>SD</i>)	Частотность (<i>ipm</i>)	Длина в слогах	Длина в фонемах
Велосипед	4.52 (.90)	1.66 (.65)	23.10	4	9
Корова	4.19 (1.07)	1.20 (.40)	44.80	3	6

Примечание: *ipm* — вхождений на миллион слов.

Таблица 2. Параметры стимулов для задания на сопоставление слова с картинкой

Существительное	Знакомство с объектом: среднее (<i>SD</i>)	Возраст усвоения: среднее (<i>SD</i>)	Частотность (<i>ipm</i>)	Длина в слогах	Длина в фонемах
Яблоко	4.80 (.64)	1.07 (.26)	53.00	3	7
Пирог	4.37 (.86)	1.47 (.67)	16.30	2	5

Примечание: *ipm* — вхождений на миллион слов.

A2. Соотнесение услышанного слова и изображения. Это задание оценивает способность пациента понимать значение услышанных слов. Испытуемому предъявляются две фотографии. На одной изображено яблоко, на другой — пирог. Пациент смотрит на изображения и слышит инструкцию: «Дотроньтесь до яблока». В данном случае слово «пирог» выступает семантическим дистрактором к целевому слову «яблоко». Оба слова характеризует фонологическая простота, что отражено в таблице 2.

A3–A4. Понимание предложения. Задания этой группы позволяют оценить чувствительность пациента к корректности предложений (на слух и визуально), а также возможность соотнести их с предъявленным изображением. Исследователь показывает пациенту изображение, над которым написано предложение, и читает предложение вслух. Испытуемый должен оценить, соответствует ли предложение изображению, и ответить «да» или «нет».

В задании A3 испытуемый видит фотографию мужчины, спускающегося с лестницы, и слышит не подходящее к ней предложение (см. пример 1). В этом примере вместо корректного употребления в родительном падеже (см. пример 2) слово «лестница» употреблено в творительном падеже. Исследователь спрашивает, соответствует ли предложение фотографии. Правильный ответ: «нет».

- (1) Мужчина спускается с лестницей.
- (2) Мужчина спускается с лестницы.

В задании A4 испытуемый видит фотографию женщины, собирающейся помыть машину. В то же время исследователь предъявляет написанное предложение со

сложным сказуемым «Женщина собирается помыть машину» и одновременно читает его вслух. Далее испытуемому задается вопрос о том, соответствует ли написанное предложение фотографии. В этом случае правильный ответ на вопрос: «да».

A5–A7. Задания, предполагающие ответ «да» или «нет». В заданиях этой группы оценивается сохранность семантического языкового уровня у пациентов с нарушениями речи. Исследователь задает вопросы, на которые испытуемый должен ответить «да» или «нет» (подробнее см. таблицу 3).

Часть А: оценка результатов

Выполнение каждого задания в части А оценивается в 1 балл, если испытуемый дает правильный ответ. В случае неправильного ответа, медленных ответов или самокоррекции испытуемый получает 0 баллов.

Часть Б: порождение речи

Часть Б состоит из семи заданий, оценивающих способность пациента к порождению речи. В задания этой группы входят называние предметов, повторение слов и чтение слов вслух. Там, где это было возможно, мы сохранили оригинальные материалы и перевели их на русский язык, а в ряде случаев заменили их на более релевантные стимулы.

B1–B3. Называние предметов. Пациенту поочередно предъявляются два изображения предметов и одно — животного. Задача состоит в том, чтобы назвать изображенные объекты. Сложность увеличивается от стимула к стимулу (см. таблицу 4). Мы заменили изображение в задании B1 (фортепьяно), чтобы избежать неоднозначности (к этому стимулу подходят ответы «пианино» и «фортепьяно»). Вместо него мы используем фотографию гитары. Также мы заменили

Таблица 3. Параметры стимулов для заданий, предполагающих ответ «да» или «нет»

Задание	Вопрос	Ожидаемый ответ	Параметры
A5	У стола есть окна?	Нет	Понимание предложений с фокусом на семантике существительного
A6	Можно ли съесть макароны?	Да	Понимание предложения с фокусом на отношениях между глаголом и существительным
A7	Бабочка — это птица?	Нет	Понимание предложений с фокусом на семантике категорий

Таблица 4. Параметры стимулов для задания на называние объектов

Существительное	Знакомство с объектом: среднее (SD)	Возраст усвоения: среднее (SD)	Частотность (ipm)	Длина в слогах	Длина в фонемах	Кол-во кластеров согласных
B1. Гитара	4.06 (1.20)	1.98 (.77)	22.20	3	6	0
B2. Черепаха	3.76 (1.33)	1.63 (.73)	6.30	4	8	0
B3. Вертолет	3.55 (1.36)	2.01 (.69)	38.70	3	8	1

стимул «брокколи» из задания Б2 в связи с его низкой частотностью в русском языке (0.7 ірm). В АБВ мы используем слово «черепаха».

Б4–Б5. Повторение слов. В этом задании пациент должен повторить два абстрактных фонологически сложных существительных (см. таблицу 5). Мы сохранили принцип оригинальной нидерландской версии, и стимул, использующийся в Б4, был переведен на русский. В задании Б5 мы выбрали слово, семантически и фонологически близкое к нидерландскому (*sollicitatie* «заявление»).

Б6–Б7. Чтение вслух. Задания Б6–Б7 позволяют оценить способность пациента читать вслух фонологически сложные слова. Два стимула выбраны по тому же принципу, что и в оригинальном тесте АВС (подробнее см. таблицу 6).

Часть Б: оценка результатов

Все задания в части Б оцениваются 1 баллом, если ответ правильный. Семантические и фонематические парафазии, «словесный салат», персеверации, ответ «я не знаю» и отсутствие ответа оцениваются в 0 баллов.

Часть В. Дополнительные задания

Часть В — дополнительная часть теста АБВ, которая проводится, если исследователь хочет узнать больше о речевом дефиците пациента. Часть В состоит из семи заданий повышенной сложности.

В1. Описание изображения. В этом задании исследователь показывает испытуемому изображение и задает вопрос: «Опишите, что не так на этом рисунке». Мы выделили пять элементов изображения, которые испытуемые без неврологических нарушений определяли как странные или неправильные: зонтик (вывернутый наизнанку), утка, штаны, облако, ботинки. Пациенты получают 1 балл, если описывают зонтик и как минимум два из четырех других элементов. Если пациент описывает в сумме меньше трех элементов изображения или не упоминает зонтик, он/она получает 0 баллов. Это задание оценивает способность к порождению нарратива.

В2. Задания, предполагающие ответ «да» или «нет». Задание содержит три вопроса, на которые нужно ответить «да» или «нет» (см. таблицу 7).

В3. Повторение словосочетаний. В этом задании пациент должен повторить три фонологически сложных словосочетания (см. таблицу 8).

Таблица 5. Параметры стимулов для задания на повторение

Существительное	Частотность (ірm)	Длина в слогах	Длина в фонемах	Кол-во кластеров согласных	Кол-во артикуляционных переключений
Б4. Организация	312.4	6	11	1	2
Б5. Электричество	20.2	4	13	2	3

Примечание: ірm — вхождений на миллион слов.

Таблица 6. Параметры стимулов для чтения вслух

Существительное	Частотность (ірm)	Длина в слогах	Длина в фонемах	Кол-во кластеров согласных	Кол-во артикуляционных переключений
Б6. Администрация	98.3	6	13	2	2
Б7. Поздравление	10.8	5	12	1	3

Примечание: ірm — вхождений на миллион слов.

Таблица 7. Параметры стимулов для заданий, предполагающих ответ «да» или «нет» (часть В)

Задание	Вопрос	Ожидаемый ответ	Параметры
В2а	Можно ли дуть в барабан?	Нет	Понимание предложения с фокусом на отношениях между глаголом и существительным
В2б	Китай — это страна?	Да	Понимание предложений с фокусом на семантике (знаниях о мире)
В2в	У слона есть клюв?	Нет	Понимание предложений с фокусом на семантике существительного

В4. Чтение вслух. В этом задании пациент должен прочитать вслух два фонологически сложных слова и одно словосочетание (см. таблицу 9).

В5. Завершение серии слов. Это задание предназначено для оценки сохранности у пациента лексического доступа. Испытуемому предъявляются три серии по три слова; каждая из серий представляет отдельную семантическую категорию: а) ромашка, тюльпан, лилия; б) щука, сельдь, лосось; в) капуста, огурец, морковь. Исследователь зачитывает каждую серию вслух и просит испытуемого ее закончить. Если пациент называет еще одного представителя той же категории или само название категории, он/она получает 1 балл. Во всех остальных случаях ответ оценивается нулем баллов.

В6. Задание на вербальную беглость. Это задание оценивает способность извлекать слова конкретной семантической категории. Пациент должен по усмотрению исследователя назвать представителей одной или нескольких категорий, а именно: «животные», «профессии», «города». Исследователь включает таймер и просит испытуемого назвать как можно больше представителей той или иной категории за одну минуту. Порог правильного выполнения задания для каждой категории был рассчитан как количество представителей категории, названное здоровым испытуемым, минус одно стандартное отклонение. Ответ оценивается 1 баллом, если пациент называет 15 и более животных; 12 и более профессий; 15 и более городов. Если количество названных представителей ниже порогового значения, испытуемый получает 0 баллов.

В7. Повторение слогов. Задание позволяет определить наличие нарушений артикуляции. Исследователь просит испытуемого повторить три серии слогов с взрывными согласными, различающимися местом образования: «папапапа» (с лабиальными соглас-

ными), «тататата» (с альвеолярными согласными) и «какакака» (с велярными согласными). При произнесении четвертой серии слогов наблюдаются артикуляторные переключения между этими согласными: «патака патака патака». Каждая правильно воспроизведенная последовательность оценивается 1 баллом, во всех остальных случаях ставится 0 баллов. Максимально за это задание можно получить 4 балла.

Исследование 1

Исследование 1 было направлено на стандартизацию теста АБВ и оценку его соответствия современным психометрическим характеристикам. Мы стандартизировали ядро теста, состоящее из частей А (понимание речи) и Б (порождение речи). Часть В не обязательна для определения наличия речевых нарушений: она содержит ряд дополнительных заданий, которые специалист по желанию может использовать для лучшего понимания речевого дефицита пациента. Поэтому часть В не была включена в Исследование 1.

Методы

Участники. В исследовании приняли участие две группы испытуемых. Клиническая группа ($N=80$, 30 женщин, средний возраст 60 лет, $SD=11.5$, размах 26–79; среднее количество лет образования = 14, $SD=2.5$, размах 8–19) состояла из людей в хроническом постинсультном периоде (в среднем 56 месяцев с момента инсульта, $SD=66$, размах 5–307). При поступлении в Центр патологии речи и нейрореабилитации (Москва, Россия) эти пациенты были обследованы штатными логопедами с помощью стандартных тестов. По результатам обследования у них были выявлены разные формы афазии (у 39 была диагностирована

Таблица 8. Параметры стимулов для задания «Повторение» (часть В)

Задание	Выражение	Длина в слогах	Длина в фонемах	Кол-во кластеров согласных	Кол-во артикуляционных переключений
V3a	Футляр для дрели	5	14	3	0
V3б	Дорожное ограждение	9	19	3	3
V3в	Опубликовать опровержение	11	24	3	4

Таблица 9. Параметры стимулов для задания «Повторение» (часть В)

Задание	Существительное / выражение	Длина в слогах	Длина в фонемах	Кол-во кластеров согласных	Кол-во артикуляционных переключений
V4a	Абонемент	4	9	1	3
V4б	Лаборатория	6	11	0	4
V4в	Невероятная конструкция	10	22	3	6

афазия беглого типа, у 21 — небеглого), а у 20 была диагностирована дизартрия без афазии. Никто из испытуемых не страдал нейродегенеративными заболеваниями и не имел истории психических и/или неврологических расстройств.

Контрольная группа ($N=120$, 61 женщина, средний возраст 43, $SD=19$, размах 17–86; среднее количество лет образования — 15, $SD=2.6$, размах 10–21) состояла из добровольцев, согласившихся на участие в исследовании. Участники контрольной группы подбирались таким образом, чтобы клиническая группа не отличалась от контрольной по возрасту, полу и уровню образования.

Все участники (преморбидно) были правшами, кроме одного левши и одного амбидекстра в клинической группе и трех левшей в контрольной группе. Участники были монолингвальными носителями русского языка, имели нормальное или скорректированное до нормального зрение и слух, не получали травм головного мозга в течение последних трех лет. Никто из испытуемых на момент исследования не принимал медикаменты, которые могли бы повлиять на их когнитивные функции. Подробная демографическая информация об испытуемых представлена в [дополнительных онлайн-материалах](#), таблица 1 (клиническая группа) и таблица 2 (контрольная групп). Все участники подписали форму информированного согласия, одобренную Комиссией по внутриуниверситетским опросам и этической оценке эмпирических исследовательских проектов НИУ ВШЭ.

Материалы и процедура. Все участники в рандомизированном порядке выполнили задания теста АБВ и планшетной версии Токен-теста (Token Test App: Bastiaanse et al., 2020; русскоязычная версия: Akinina, Buiivolova, Soloukhina, Bastiaanse, 2019). Токен-тест — тест для выявления речевых нарушений (De Renzi, Faglioni, 1978). В процессе тестирования участник видит разноцветные фигуры (квадраты и круги разных цветов и размеров) и слышит инструкции, например: «Дотроньтесь до черного квадрата», «Дотроньтесь до желтого круга и до зеленого квадрата», «Дотроньтесь до всех кругов, кроме зеленого». Тест состоит из пяти серий заданий, уровень сложности которых возрастает по мере выполнения. Оба теста проводились клиническим лингвистом в тихом помещении. Ответы на задания теста АБВ записывались на диктофон после получения дополнительного устного согласия, после чего исследователь заносил ответы в бумажный протокол и оценивал их. Ответы, полученные в результате выполнения Токен-теста, оценивались в приложении автоматически.

Чтобы определить уровень *надежности* теста АБВ, мы оценивали его чувствительность, специфичность, положительную и отрицательную прогностическую значимость, отношение правдоподобия положительного и отрицательного результата. Чувствительный тест позволяет верно определить наличие у испытуемого афазии. Специфичность определяется тем, насколько точно тест исключает людей без афазии. Положительная и отрицательная прогностическая значимость — мера вероятности того, что при положительном/отрицательном результате теста испы-

туемый имеет / не имеет афазии. Отношение правдоподобия положительного и отрицательного результата оценивает вероятность положительного/отрицательного результата для человека с афазией или без нее. Все перечисленные параметры оценивались на основании формулы, предложенной в работе Т. Гринхальх (Greenhalgh, 1997).

Мы оценили *конкурентную валидность* теста АБВ (Ivanova, Hallowell, 2013) с помощью корреляции между результатами, полученными при выполнении АБВ и Токен-теста, достоверность и надежность которого доказана (Akinina, Bastiaanse, Soloukhina, 2019; Akinina, Buiivolova, Soloukhina, Bastiaanse, 2019). *Конструктивная валидность* оценивалась с помощью сопоставления результатов выполнения АБВ и степени выраженности речевого нарушения на основании логопедического заключения.

Чтобы измерить *межэкспертную надежность*, четыре эксперта независимо друг от друга оценили записанные ответы 30 участников (с афазией и/или дизартрией), вслепую выбранных из клинической группы.

Для оценки *внутренней согласованности* теста мы измерили, как каждое задание коррелирует с итоговым баллом, по формуле Кудера-Ричардсона (KR-20). Мы присвоили заданиям уровень сложности, чтобы удостовериться, что тест содержит задания разной сложности. Мы измерили процент правильных ответов для каждого задания в репрезентативной выборке людей с афазией разных типов и степени выраженности (всего 60 человек с афазией из клинической группы) и сравнили с результатами неврологически здоровых участников. Кроме того, мы оценили, в какой степени каждый из стимулов вносит вклад в финальную оценку.

Чтобы оценить влияние демографических характеристик на финальный результат, мы сопоставили ответы, полученные при выполнении АБВ с возрастом и уровнем образования испытуемых в клинической и контрольной группе, а также провели корреляционный анализ между результатами по АБВ и сроком давности заболевания в клинической группе. Помимо этого, мы сравнили результаты по тесту АБВ в группах участников с разными типами афазии (беглой vs небеглой).

Результаты

Средний балл, полученный участниками за выполнение теста АБВ в клинической группе, составил 10.50 ($SD=3.29$, размах 2–14), средний балл по Токен-тесту был 20.16 ($SD=9.02$, размах 0.50–36). В группе участников с дизартрией ($N=20$) средний балл за выполнение теста АБВ составил 12.5 балла ($SD=2.08$, размах 7–14), а средний балл по Токен-тесту составил 29.05 ($SD=7.86$, размах 7.50–36). В контрольной группе участники за выполнение заданий теста АБВ получили в среднем 13.54 балла ($SD=0.74$, размах 8–14), а за выполнение Токен-теста — 34.55 балла ($SD=1.90$, размах 27–36).

Мы определили пороговое значение, позволяющее разделить между собой результаты людей с речевыми нарушениями и без них, как равное значению 5-й перцентили результатов контрольной группы. Для основной части теста, включающей части А и Б, пороговое значение составило 13 из 14 баллов. Такой же

порог был установлен для оригинальной нидерландской версии теста.

Данные о чувствительности, специфичности, положительной и отрицательной прогностической значимости, отношении правдоподобия положительного и отрицательного результата, полученные по формуле Т. Гринхалх (Greenhalgh, 1997), приведены в таблице 10. Для этого результаты выполнения АБВ сравнивались с результатами Токен-теста, принятого в качестве эталона. Мы включили результаты, полученные в контрольной группе и в группе людей с диагностированной логопедами афазией, но не дизартрией, потому что Токен-тест, в отличие от теста АБВ, не чувствителен к нарушениям артикуляции. Исходя из этого, результаты пациентов с дизартрией были исключены из анализа во избежание неоднозначности. Результаты представлены в таблице 10.

Результаты демонстрируют, что тест АБВ может считаться достаточно чувствительным, поскольку с помощью него мы смогли верно определить наличие афазии у 72 % участников, а также высоко специфичным, так как смогли верно определить отсутствие афазии у 96 % участников, не имеющих ее. Тест имеет высокую положительную (.88) и отрицательную (.89) прогностическую значимость, то есть вероятность, что люди с афазией получают положительный результат, а люди без афазии — отрицательный. Отношение правдоподобия положительного результата теста (17.35) указывает на очень высокую вероятность положи-

тельного результата у человека с афазией, а отношение правдоподобия отрицательного результата (0.29) говорит о средней вероятности отрицательного результата для человека без афазии (McGee, 2002).

Мы измерили наличие и оценили значимость корреляции между результатами, полученными при выполнении теста АБВ и Токен-теста в трех группах испытуемых. В контрольной группе коэффициент корреляции Пирсона был низким, но значимым ($r = .356, p < .001$). В группе пациентов с афазией корреляция Пирсона была умеренной и значимой ($r = .592, p < .001$). В группе людей с дизартрией результаты выполнения АБВ- и Токен-теста не показали значимой корреляции ($r = -.109, p = .645$), что объясняется тем, что Токен-тест, в отличие от АБВ, не предназначен для выявления расстройств артикуляции (например, дизартрии). Значимая корреляция между результатами обоих тестов для группы людей с афазией доказывает, что АБВ так же эффективен для определения афазии, как и Токен-тест.

Чтобы определить конструктивную валидность АБВ, мы рассчитали коэффициент корреляции Пирсона между результатами выполнения АБВ и степенью выраженности речевых нарушений (от легкой до очень грубой по шкале от 1 до 6), диагностированной логопедами с помощью стандартных тестов. Мы обнаружили умеренную значимую корреляцию ($r = -.481, p < .001$), что говорит о возможности обнаружить речевые нарушения и оценить их грубость с помощью АБВ.

Таблица 10. Надежность АБВ

	ТТ < 29 (n = 53)	ТТ ≥ 29 (n = 121)
АБВ < 13 (n = 42)	A Истинно положительный (38)	B Ложноположительный (5)
АБВ ≥ 13 (n = 132)	C Ложноотрицательный (15)	D Истинно отрицательный (116)

	Параметры	
	Значение	95 % доверительный интервал
Чувствительность $A/(A+C)$.72	.58 – .83
Специфичность $D/(B+D)$.96	.91 – .99
Положительная прогностическая значимость $A/(A+B)$.88	.75 – .96
Отрицательная прогностическая значимость $D/(C+D)$.89	.82 – .93
Отношение правдоподобия положительного результата чувствительность / (1-специфичность)	17.35	7.23 – 41.61
Отношение правдоподобия отрицательного результата (1-чувствительность) / специфичность	.30	.19 – .45

Примечание: АБВ < 13: оценка по АБВ ниже 13, что указывает на наличие речевых нарушений. АБВ ≥ 13: оценка по АБВ 13 или выше, что указывает на отсутствие речевых нарушений. ТТ < 29: балл за выполнение Токен-теста ниже 29, что указывает на наличие афазии. ТТ ≥ 29: оценка Токен-теста равна или выше 29, что указывает на отсутствие афазии.

Коэффициент корреляции отрицательный, поскольку тесты имеют разнонаправленные шкалы.

Для определения межэкспертной надежности четверо экспертов (двое нейролингвистов и двое студентов лингвистических направлений НИУ ВШЭ) независимо друг от друга оценили записи ответов 30 испытуемых, случайно выбранных из клинической группы. Эксперты оценивали задания частей А и Б (за исключением заданий А1–А2, где оценка на основании аудиозаписи невозможна, так как ответ испытуемого заключается в выборе правильного изображения). После исключения этих двух заданий межэкспертная надежность составила 12 баллов с максимальным результатом в 5 баллов по части А и 7 — по части Б. Средний балл оценки составил 9.90 ($SD=2.41$, размах 3–12) для Эксперта 1; 9.78 ($SD=2.21$, размах 3–12) для Эксперта 2; 9.77 ($SD=2.29$, размах 4–12) для Эксперта 3; и 10 ($SD=2.18$, размах 3–12) для Эксперта 4. Чтобы оценить согласованность оценок между экспертами, мы рассчитали внутриклассовый коэффициент корреляции (двусторонний случайный, средний балл ICC) (ICC; McGraw, Wong, 1996). Полученный внутриклассовый коэффициент корреляции, равный .955 (95% CI:

.922–.977; $F(29, 89.4) = 22.1$, $p < .001$), может считаться очень высоким (Коо, Ли, 2016).

Внутренняя согласованность теста была измерена с помощью формулы Кудера-Ричардсона (KR-20), поскольку данные имеют дихотомический характер (Cicchetti, 1994; Ivanova, Hallowell, 2013). В соответствии с результатами выполнения АБВ, тест может считаться гомогенным (KR-20 = .850) (Patock, 2004). Сравнение стимулов между собой показало, что они сопоставимы по сложности: пять заданий могут считаться легкими, семь — умеренно сложными и одно — сложным (Office of Educational Assessment, 2020). Степень соотношенности каждого задания с финальным баллом варьируется от слабой до хорошей (Office of Educational Assessment, 2020). Подробный анализ стимулов представлен в таблице 11.

Чтобы выявить влияние демографических факторов на результаты выполнения теста АБВ, мы рассчитали корреляцию Пирсона между а) возрастом участника и результатами выполнения АБВ в контрольной ($r = -.142$, $p = .120$) и клинической ($r = -.113$, $p = .320$) группах; б) уровнем образования участника и результатами выполнения АБВ в контрольной ($r = .145$, $p = .105$)

Таблица 11. Сложность стимулов

Задание	Сложность: среднее (SD)		Сложность: интерпретация значения	Различимость	Различимость: интерпретация значения
	NBD	PWA			
A1	.99 (.09)	.93 (.25)	Легко	.15	Достаточная
A2	1.00 (.00)	.97 (.18)	Легко	.10	Достаточная
A3	.73 (.47)	.37 (.49)	Сложно	.55	Хорошая
A4	.95 (.24)	.92 (.28)	Легко	.10	Достаточная
A5	.98 (.13)	.82 (.39)	Умеренно	.35	Хорошая
A6	.99 (.09)	.87 (.34)	Легко	.25	Достаточная
A7	.99 (.09)	.93 (.25)	Легко	.20	Достаточная
B1	.99 (.09)	.78 (.42)	Умеренно	.60	Хорошая
B2	1.00 (.00)	.67 (.48)	Умеренно	.50	Хорошая
B3	.98 (.13)	.70 (.46)	Умеренно	.75	Хорошая
B4	1.00 (.00)	.67 (.48)	Умеренно	.90	Хорошая
B5	1.00 (.00)	.65 (.48)	Умеренно	.90	Хорошая
B6	1.00 (.00)	.67 (.48)	Умеренно	.85	Хорошая
B7	.99 (.09)	.57 (.50)	Умеренно	.75	Хорошая

Примечание: NBD — группа испытуемых без неврологических нарушений, PWA — группа людей с афазией.

и клинической ($r = .106$, $p = .349$) группах; в) сроком давности заболевания и результатами выполнения АБВ в клинической группе ($r = -.171$, $p = .127$). Мы также разделили контрольную группу на две подгруппы по возрасту (17–45 лет, $n = 64$; и 45–86 лет, $n = 56$) и не выявили различия в их результатах (критерий Вилкоксона: $Z = -1.403$, $p = .161$). Две возрастные подгруппы участников клинической группы (26–59 лет, $n = 34$; и 60–86 лет, $n = 46$) также показали схожие результаты ($Z = -0.238$, $p = .778$). Наконец, мы не обнаружили различия в результатах пациентов с беглыми ($n = 38$) и небеглыми ($n = 22$) типами афазии ($Z = -1.036$, $p = .300$). Как и ожидалось, значимо отличались результаты групп людей с афазией и с дизартрией ($Z = -3.296$, $p = .001$). В группе людей с дизартрией не было обнаружено значимых различий в результатах выполнения заданий, направленных на оценку понимания и порождения речи ($Z = -.524$, $p = .600$).

Промежуточное обсуждение

В исследовании 1 мы показали, что скрининговый тест АБВ для диагностики речевых нарушений соответствует современным психометрическим стандартам. Этот тест может считаться надежным, так как такие параметры, как чувствительность, специфичность, положительная и отрицательная прогностическая значимость, высоки, а отношения правдоподобия положительного и отрицательного результата тестов имеют приемлемые значения. Мы обнаружили умеренную значимую корреляцию между результатами теста АБВ и Токен-теста в группе людей с афазией, что свидетельствует о том, что АБВ позволяет зафиксировать наличие афазии так же хорошо, как и Токен-тест. Конструктивная валидность АБВ была оценена как умеренная, поскольку была обнаружена умеренная значимая корреляция между результатами выполнения теста АБВ и результатами выполнения стандартных тестов, используемых логопедами в клиниках. Четыре независимых эксперта оценили результаты выполнения теста участниками, вслепую выбранными из клинической группы. Межэкспертная надежность, определяющая степень согласованности оценок, была оценена как высокая. Это значит, что вне зависимости от того, кто будет оценивать результаты тестирования, эти результаты могут считаться достоверными.

Анализ стимулов теста показал, что тест может считаться гомогенным, а сами стимулы варьируются по сложности (легкая, умеренная, высокая). Тестирование контрольной группы показало, что люди без патологий мозга выполняют задания теста практически безошибочно (95–100 % правильных ответов), при этом больше всего затруднений вызвало выполнение задания А3. Однако результат выполнения этого задания был значительно выше среднего (73 % правильных ответов), поэтому мы решили не менять стимульный материал для этого задания. Кроме того, результат анализа стимулов показал, что стимулы с умеренной и высокой сложностью более значимо коррелируют с общим результатом и играют большую роль при определении степени выраженности речевого нарушения. Принимая во внимание полученные результаты, мы считаем, что финальный набор стимулов те-

ста АБВ, включающий в себя пять стимулов с низкой, семь — с умеренной сложностью и один стимул с высокой сложностью, является достаточным для различения пациентов с разной степенью выраженности речевых нарушений (Ivanova, Hallowell, 2013).

Наконец, мы выяснили, что демографические характеристики не оказывают влияния на результат выполнения теста АБВ. Помимо этого, не было выявлено значимых различий между результатами участников с разными типами афазии, что значит, что вне зависимости от типа афазии тест может достоверно зафиксировать наличие такого нарушения. Тем не менее было зафиксировано значимое различие между результатами выполнения теста в группах людей с афазией и с дизартрией. Общий балл АБВ в группе людей с дизартрией без афазии был значимо выше, чем в группе людей с афазией, хотя и ниже порогового значения. Кроме того, в этой группе не было отмечено значимой разницы между результатами выполнения заданий на понимание и заданий на порождение речи. В то же время результаты выполнения Токен-теста в группе людей с дизартрией были выше порогового значения. Таким образом, мы не можем утверждать, что АБВ позволяет различить наличие афазии и наличие артикуляторных нарушений. Однако комбинация АБВ и Токен-теста может быть полезной при составлении программы реабилитации.

Исследование 2

После стандартизации теста в популяции людей со стабильным речевым статусом мы исследовали, подходит ли АБВ для использования в остром постинсультном периоде. Мы оценили, насколько результаты выполнения теста АБВ могут быть сопоставимы с результатами выполнения другого скринингового теста — Быстрого теста на афазию, который был ранее стандартизирован (Buiivolova et al., 2020).

Методы

Участники. В клиническую группу вошли 20 человек (7 женщин; средний возраст — 61.2 года, $SD = 13.9$, размах 32–85). Все участники перенесли инсульт (среднее время от начала заболевания — 10 дней, $SD = 4.8$, размах 2–23 дня), нейродегенеративных расстройств у испытуемых в анамнезе не было. На момент исследования участники находились на лечении в неврологическом отделении 31-й Московской городской клинической больницы. Согласно оценке логопеда, у 10 участников была зафиксирована афазия, у 6 участников — дизартрия, а у 4 участников не было отмечено речевых нарушений. Демографические данные участников представлены в [дополнительных онлайн-материалах](#), таблица 3.

Материалы и процедура. Все участники были протестированы клиническим лингвистом или логопедом Быстрым тестом на афазию и основной частью АБВ (части А и Б) в рандомизированном порядке. БТА был выбран для исследования, поскольку это валидный и стандартизированный скрининговый тест для обнаружения речевых нарушений, следовательно, он

может считаться «золотым стандартом». Он состоит из 6 заданий, по которым оценивается понимание инструкций, повторение трех слов и одного предложения, называние трех объектов по рисунку, оценки дизартрии и задания на вербальную беглость. В тот же день логопед предоставлял подробную речевую оценку по шкале Л. И. Вассермана для оценки степени выраженности речевых нарушений у больных с локальными поражениями мозга (Вассерман и др., 1997). По итогам комплексного обследования логопед определял степень выраженности речевых нарушений от 1 (легкая) до 6 (очень грубая). Все участники или их родственники подписали информированное согласие и согласие на обработку персональных данных при поступлении в больницу. Экспериментатор приступал к тестированию пациентов после объяснения процедуры и получения устного согласия на участие в исследовании.

Результаты

Средний балл за выполнение БТА составил 7.60 при максимально возможных 26 баллах ($SD=7.50$, размах 0–23), за выполнение АБВ — 10.25 из 14 возможных баллов ($SD=3.58$, размах 2–14). Коэффициент корреляции (Спирмена) между результатами АБВ и БТА был высоким и значимым ($\rho = -.911$, $p < .001$). Результаты выполнения теста АБВ также имели высокую значимую корреляцию со степенью выраженности речевого нарушения, согласно оценке логопеда ($\rho = -.835$, $p < .001$). Коэффициенты корреляции отрицательные, поскольку тесты имеют разнонаправленные шкалы. Также мы обнаружили высокую значимую корреляцию между результатами выполнения теста БТА и оценкой логопеда ($\rho = .856$, $p < .001$).

В группе людей с афазией и/или дизартрией средний балл за выполнение АБВ составил 9.50 баллов ($SD=3.63$, размах 2–14), в то время как люди, перенесшие инсульт без каких-либо сопутствующих речевых нарушений, выполняли тест АБВ выше порогового значения (13 баллов) со средним баллом 13.25 ($SD=0.50$, размах 13–14). Результаты выполнения в этих двух группах значимо различались (критерий Вилкоксона: $Z = -3.308$, $p < .001$).

Промежуточное обсуждение

Согласно результатам, полученным в ходе второго исследования, тест АБВ так же подходит для скрининга речи, как и русскоязычная версия БТА. Они измеряют один и тот же конструкт, а результаты выполнения обоих тестов значимо коррелируют с результатами оценки логопеда. Несмотря на то, что оба теста могут быть успешно использованы для скрининга речевых патологий, АБВ позволяет более детально оценить состояние речевой функции пациента (например, АБВ позволяет оценить способность к пониманию предложений и чтению вслух) в течение периода времени, сопоставимого с тем, который может быть затрачен на выполнение заданий БТА. Кроме того, АБВ позволяет независимо друг от друга оценить понимание и порождение речи, а при необходимости использовать дополнительные задания для более детальной оценки речевого дефицита. Результаты исследования показывают, что с помощью

теста АБВ возможно сделать различие между пациентами, имеющими речевые нарушения, и теми, у кого таких нарушений не отмечалось.

Обсуждение

Мы провели два исследования, чтобы оценить, соответствует ли русскоязычная версия теста «Афазия: быстрая верификация» современным психометрическим стандартам и может ли она использоваться для скрининга речевых патологий в остром постинсультном периоде. АБВ позволяет выявлять проблемы на уровне понимания и порождения речи, а также включает дополнительные задания, которые могут быть информативными для разработки дальнейшего курса реабилитации для каждого пациента. Проведение теста не требует каких-либо специальных навыков и может выполняться не только логопедами или нейропсихологами, но также неврологами и другими сотрудниками неврологического отделения. Результаты исследования показывают, что тест соответствует стандартам и может использоваться в остром постинсультном периоде.

Наше исследование имеет некоторые ограничения. Во-первых, из-за пандемии COVID-19 мы не смогли собрать данные для анализа ретестовой надежности, чтобы понять, стабильны ли результаты выполнения теста с течением времени. Однако для АБВ это не критично, поскольку этот тест предназначен для скрининга, а не для оценки динамики восстановления речевой функции. Известно, что в острой стадии инсульта высока вероятность спонтанного восстановления (Holland, Fridriksson, 2001), поэтому неясно, будут ли вызваны изменения в результатах низкой ретестовой надежностью теста или процессами спонтанного восстановления. Чтобы оценить, насколько высока ретестовая надежность АБВ, необходимо будет сравнить результаты выполнения теста в группе людей с речевыми нарушениями в остром периоде и людей с хроническими речевыми нарушениями и оценить, есть ли существенные различия между показателями в двух временных точках в этих двух группах. В данный момент это — задача для будущего исследования.

Вторым ограничением является небольшое количество участников в исследовании 2. Однако в исследовании 1 участвовало 120 неврологически здоровых испытуемых, 80 людей с хронической афазией и 20 людей с хронической дизартрией. В исследовании 2 результаты, полученные в группе из 20 человек в остром постинсультном периоде, подтвердили результаты исследования в хронической популяции и показали высокую корреляцию с существующими диагностическими и скрининговыми инструментами. Таким образом, для текущего исследования размер выборки можно считать достаточным.

Наконец, основной целью текущего исследования была стандартизация основной части АБВ, которая включает в себя части А (понимание речи) и Б (порождение речи). Необязательная часть В (дополнительные задания) была адаптирована для русского языка, но, поскольку она является необязательной, ее стандартизация не входила в задачи настоящего исследования.

АБВ также представляется перспективным инструментом для использования в нейрохирургической практике. В настоящее время не существует быстрого и достоверного способа определения наличия афазии и тяжести речевых расстройств до и после операции, особенно у пациентов с серьезными речевыми нарушениями. Вот почему АБВ следует стандартизировать не только в группе людей, перенесших инсульт, но и в группе людей с опухолями головного мозга и с черепно-мозговыми травмами.

Таким образом, результаты стандартизации русскоязычной версии теста «Афазия: быстрая верификация» можно считать многообещающими. Этот тест можно рассматривать как надежный инструмент для скрининга речевых нарушений у пациентов в остром постинсультном периоде.

Литература

- Акинина Ю. С., Грабовская М. А., Вечкаева А., Игнатев Г. А., Исаев Д. Ю., Ханова А. Ф. Библиотека психолингвистических стимулов: новые данные для русского и татарского языка // Седьмая международная конференция по когнитивной науке: Тезисы докладов. Светлогорск, 20–24 июня 2016 г. / Под ред. Ю. И. Александрова, К. В. Анохина. М.: Институт психологии РАН, 2016. С. 93–95.
- Вассерман Л. И., Дорофеева С. А., Меерсон Я. А. Методы нейропсихологической диагностики: Практическое руководство. СПб.: Стройлеспечать, 1997.
- Щербатова М. М., Котов С. В. Компьютерный скрининг-тест афазий // *Consilium Medicum*. 2017. Т. 19. № 2. С. 24–26.
- Akhutina T. Luria's classification of aphasia and its theoretical basis // *Aphasiology*. 2015. Vol. 30. No. 8. P. 878–897. <https://doi.org/10.1080/02687038.2015.1070950>
- Akinina Y., Bastiaanse R., Soloukhina O. Automated language deficit diagnostics with The Token Test App: Test-retest reliability and practice effects // Paper presented on Conference on Serious Games and Applications for Health, Kyoto, Japan, 2019, August 5–7. 2019.
- Akinina Y., Buivolova O., Soloukhina O., Bastiaanse R. Psychometric properties of the Token Test App // *Stem-, Spraak- en Taalpathologie*. 2019. Vol. 24. No. Supplement. P. 74–76. URL: <https://sstp.nl/article/view/35738/33292>.
- Azuar C., Leger A., Arbizu C., Henry-Amar F., Chomel-Guil-laume S., Samson Y. The Aphasia Rapid Test: An NIHSS-like aphasia test // *Journal of Neurology*. 2013. Vol. 260. No. 8. P. 2110–2117. <https://doi.org/10.1007/s00415-013-6943-x>
- Bastiaanse R., Akinina Y., Satoer D., de Kok D. The Token Test App, Second Edition. Groningen, Netherlands: 2020. URL: <https://play.google.com/store/apps/details?id=nl.rug.getc.tokenest>.
- Buivolova O., Vinter O., Bastiaanse R., Dragoy O. The Aphasia Rapid Test: Adaptation and standardisation for Russian // *Aphasiology*. 2020. P. 1–15. <https://doi.org/10.1080/02687038.2020.1727836>
- Cicchetti D. V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology // *Psychological Assessment*. 1994. Vol. 6. No. 4. P. 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- De Renzi E., Faglioni P. Normative data and screening power of a shortened version of the Token Test // *Cortex*. 1978. Vol. 14. No. 1. P. 41–49. [https://doi.org/10.1016/s0010-9452\(78\)80006-9](https://doi.org/10.1016/s0010-9452(78)80006-9)
- El Hachoui H., Visch-Brink E. G., de Lau L. M. L., van de Sandt-Koenderman M. W. M. E., Nouwens F., Koudstaal P. J., Dip-pel D. W. J. Screening tests for aphasia in patients with stroke: A systematic review // *Journal of Neurology*. 2016. Vol. 264. No. 2. P. 211–220. <https://doi.org/10.1007/s00415-016-8170-8>
- Greenhalgh T. How to read a paper: Papers that report diagnostic or screening tests // *British Medical Journal*. 1997. Vol. 315. No. 7107. P. 540–543. <https://doi.org/10.1136/bmj.315.7107.540>
- Holland A., Fridriksson J. Aphasia management during the early phases of recovery following stroke // *American Journal of Speech-Language Pathology*. 2001. Vol. 10. No. 1. P. 19–28. [https://doi.org/10.1044/1058-0360\(2001\)004](https://doi.org/10.1044/1058-0360(2001)004)
- Institute for Health Metrics and Evaluation (IHME). Russia profile. Seattle, WA: IHME, University of Washington, 2018. URL: <http://www.healthdata.org/Russia> (09.06.2020).
- Ivanova M. V., Hallowell B. A tutorial on aphasia test development in any language: Key substantive and psychometric considerations // *Aphasiology*. 2013. Vol. 27. No. 8. P. 891–920. <https://doi.org/10.1080/02687038.2013.805728>
- Koo T. K., Li M. Y. Cracking the code: Providing insight into the fundamentals of research and evidence-based practice. A guideline of selecting and reporting intraclass correlation coefficients for reliability research // *Journal of Chiropractic Medicine*. 2016. Vol. 15. No. 2. P. 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Lamers E., Ruiter M., El Hachoui H., Visch-Brink E. G. Diagnostiek in de acute fase van een beroerte. Een eerste normering van de Afasie Bedside Check (ABC) // *Nederlands Tijdschrift voor Logopedie*. 2020. Vol. 92. P. 20–25.
- Lucot C., Koleck M., Laurent K., Darrigrand B., Bordes J., Joseph P.-A., Dehail P., Mazaux J.-M. Quality of life in stroke patients with aphasia // *Annals of Physical and Rehabilitation Medicine*. 2013. Vol. 56. P. e46. <https://doi.org/10.1016/j.rehab.2013.07.018>
- McGee S. Simplifying likelihood ratios // *Journal of General Internal Medicine*. 2002. Vol. 17. No. 8. P. 647–650. <https://doi.org/10.1046/j.1525-1497.2002.10750.x>
- McGraw K. O., Wong S. P. Forming inferences about some intraclass correlation coefficients // *Psychological Methods*. 1996. Vol. 1. No. 1. P. 30–46. <https://doi.org/10.1037/1082-989x.1.1.30>
- National Institute of Neurological Disorders and Stroke (U.S.). NIH stroke scale. Bethesda, Md.: National Institute of Neurological Disorders and Stroke, Dept. of Health and Human Services, USA, 2011. URL: https://www.ninds.nih.gov/sites/default/files/NIH_Stroke_Scale_Booklet.pdf.
- National Stroke Foundation Clinical guidelines for stroke management 2010. Melbourne, Australia: 2010.
- Office of Educational Assessment. Understanding item analyses. Seattle, WA: University of Washington, 2020. URL: <https://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis>.
- Patock J. Exam scores: How to interpret your statistical analysis reports University Testing Services, 2004. URL: https://uoeee.asu.edu/sites/default/files/docs/Guide_stat_analy_exam_scores.pdf.
- Rohde A., Worrall L., Godecke E., O'Halloran R., Farrell A., Massey M. Diagnosis of aphasia in stroke populations: A systematic review of language tests // *PLOS ONE*. 2018. Vol. 13. No. 3. P. e0194143. <https://doi.org/10.1371/journal.pone.0194143>
- Visch-Brink E., El Hachoui H. Afasie Bedside Check. Rotterdam, Netherlands: Erasmus MC, 2013.