APHASIOLOGY, 2009, 23 (5), 544–556

Psychology Press
Taylor & Francis Group

Short form of the Bilingual Aphasia Test in Russian: Psychometric data of persons with aphasia

Maria V. Ivanova and Brooke Hallowell
Ohio University, Athens, OH, USA

Background: There is currently a lack of standardised aphasia batteries available in the Russian language. The psychometric properties of a short form of the Russian version of the Bilingual Aphasia Test (BAT) (Paradis, 1987) were examined. The BAT (Paradis & Zeiber, 1987) is one of the few published tests in Russian.
Aims: The primary aims were: (1) to describe the psychometric properties of a modified short form of the BAT in Russian by analysing the data collected on a large sample of Russian-speaking adults with aphasia; (2) to identify needs for further modification, validation, and standardisation; and (3) to provide a preliminary evidence base for clinicians and investigators using the test.
Methods & Procedures: The modified short form of the Russian BAT was administered to 83 patients with mild to severe aphasia. All were native speakers of Russian.
Outcomes & Results: The test was effective in discriminating patients according to level of severity of language impairment. Most of the tasks constituting the short form of the test had strong internal consistency. These results support the utility of each assessed component of the BAT in quantifying language deficits in speakers of Russian with aphasia. However, problems with the internal consistency and the validity of some items were identified.
Conclusions: The study provides preliminary data on the psychometric properties of an aphasia test in Russian. Needs for modification of the test, suggestions for further development of the test, and recommendations for further study of its psychometric properties are discussed.

Keywords: Neurogenic communication disorders; Aphasia; Language assessment; Test development; Standardised testing.

At present there are no standardised aphasia assessment batteries available in Russian. Existing Russian tests of language functioning in aphasia (Cvetkova, Axytina, & Pulaeva, 1981; Luria, 1976; Paradis & Zeiber, 1987) do not have documented norms and lack published psychometric data pertaining to their

Address correspondence to: Maria I. Ivanova, School of Hearing, Speech and Language Sciences, Ohio University, Grover Center, W218, Athens, OH 45701, USA. E-mail: mi289405@ohio.edu

This work was supported in part by grant number DC00153-01A1 from the National Institute on Deafness and Other Communication Disorders, an Ohio University Graduate Fellowship in the School of Hearing, Speech and Language Sciences, and College of Health and Human Services Student Research and Scholarly Activity Award. Dr Karen Evans-Romaine provided excellent assistance in certifying Russian translations for the informed consent process. The authors thank Victor Shklovsky and the staff of the Moscow Federal Center of Speech Pathology and Neurorehabilitation for their continuous support of the project, especially Marina Emeliyanova, for assistance with patient recruitment and data collection.

http://www.psypress.com/aphasiology DOI: 10.1080/02687030701800784

reliability and validity. This shortage of psychometrically evaluated assessment instruments makes it challenging for clinicians to draw clear and consistent distinctions between normal and impaired abilities, to compare performance among patients, and to measure reliably gains made in treatment (Allen & Yen, 2002; Roberts, 2001; Spreen & Risser, 2003). It also impedes research on aphasia in general, as such research requires standardised, valid, and reliable quantification of linguistic deficits (Fishman & Galguera, 2003). The current work aims to address directly this lack of assessment tools by taking a first step in collecting psychometric data on an existing aphasia test in Russian.

The Bilingual Aphasia Test (BAT)[1] (Paradis, 1987) is a well-known published aphasia test that has been translated into many languages. Although the multiple-language versions of the test were initially designed to assess bilingual patients, any single version of the test can be used on its own to assess language functioning in a single language (Paradis, 1987).[2] The Russian version of the BAT (Paradis & Zeiber, 1987) is one of the few published tests for aphasia that exists in the Russian language. Results of a study employing this test revealed significant increases in scores consistent with clinical observations of improved language abilities in adults with aphasia following intervention (Ivanova, Hallowell, Kruse, Shklovsky, & Emeliyanova, 2007). These findings provide indirect support for the construct validity of this assessment tool. However, no norms have been reported in the literature regarding this version of the test.

The primary objectives of this study are: (1) to describe the psychometric properties of an adapted short form of the BAT in Russian by analysing the data collected on a large sample of patients with aphasia; (2) to identify needs for further modification and standardisation of the test; and (3) to provide a preliminary evidence base for clinicians and investigators using the Russian language version of the test.

METHOD

Participants

Participants with aphasia were recruited from the inpatient unit of the Moscow Federal Center of Speech Pathology and Neurorehabilitation. Patients in the unit undergo a 45-day period of treatment with intense speech, language, and cognitive therapy at a frequency of two to three times per day, regardless of age or time post-onset (Shklovsky, 2003). Aphasia in this study is defined as "an acquired communication disorder caused by brain damage, characterized by an impairment of language modalities: speaking, listening, reading, and writing; it is not the result of a sensory deficit, a general intellectual deficit, or a psychiatric disorder" (Hallowell & Chapey, in press). Diagnosis of aphasia was determined by the first author (a neuropsychologist in Russia) and the staff of the rehabilitation centre (Russian speech-language pathologists and neuropsychologists). Presence of brain damage in participants with aphasia was verified through radiology reports based on computer tomography and/or magnetic resonance imaging. Patients with pronounced

[1] The BAT is currently out of print. One can obtain copies of the test by writing directly to Dr Paradis.

[2] The study of bilingual aphasia is not a focus of this article. Literature is available on the topic of bilingual aphasia in general (cf. Roberts, 2001) and on the BAT in particular (cf. Paradis, 1987).

attention, memory, and executive functioning impairments as identified through neuropsychological evaluation (Luria, 1976) were excluded from participation.

To be included in the study individuals with aphasia were required to pass vision and hearing screenings. Visual acuity for near vision was assessed with the Lea Symbols Line test (Hyvärinen, Näsänen, & Laurinen, 1980), which contains symbols that vary in size. Participants had to be able to identify symbols presented at a visual angle of 2 degrees at a viewing distance of 2 feet to ensure that they would be able to clearly see the images in the aphasia test (see Hallowell, in press). Participants were permitted to use glasses or contact lenses. Hearing acuity was assessed at 500, 1000, and 2000 Hz at 30 dB SPL. Participants were required to pass the screening at all three frequencies.

The modified short form of the BAT (described below) was administered to a total of 83 patients (mean age 48.12 years, range: 15–74; mean education 13.94 years, range: 9–16; M:F = 62:21) as a part of three larger studies (Hallowell, Kruse, Shklovsky, Ivanova, & Emeliyanova, 2006; Heuer et al., 2006; Ivanova et al., 2007). All patients were right-handed and had one or more focal lesions to the left hemisphere, some with additional subcortical lesions. Of the participants with aphasia, 68 (81.9%) had experienced cerebrovascular accidents (ischaemic or haemorrhagic strokes), 14 (16.9%) had experienced traumatic brain injury, and 1 (1.2%) had herpes encephalitis. Of the 68 stroke patients, 7 (10.3%) had two cerebrovascular accidents each. The patients in this sample were at least 1 month post-onset (mean time post-onset 1 year 11 months, range: 1 month to 14 years 6 months). All patients were native speakers of Russian; none was bilingual from birth. All patients were born in the former USSR (where Russian was the primary language) or in Russia. Knowledge of additional languages was not controlled; only performance in native-language Russian was pertinent to the study's objectives. The phonology and grammatical structure of the Russian language tend to be consistent in much of Russia. Regional dialects were considered to have no or minimal impact on patients' performance.

The multidimensional approach to aphasia (Hallowell & Chapey, in press) was used to classify type of aphasia. A total of 44 patients (53%) had an anterior type of aphasia, including 32 (38.6%) with Broca's aphasia, and 12 (14.5%) with transcortical motor aphasia; 39 (47%) had a posterior type of aphasia, among them 9 (10.8%) Wernicke's aphasia, 25 (30.1%) transcortical sensory aphasia, 3 (3.6%) conduction aphasia, and 2 (2.4%) anomic aphasia. Diagnosis of aphasia type was based on non-standardised neuropsychological evaluation (Luria, 1976) and criterion-referenced tasks. Aphasia severity was subjectively rated by two speech-language pathologists who worked with the patient on a daily basis at the centre. Patients were rated as having mild (27%), moderate (41%), severe (25%), or very severe aphasia (7%), depending on their receptive and expressive language abilities.

Modification of the test materials

Paradis (1987) states that the short form of the BAT may be given when time does not permit administration of the whole test. In the present study the short form of the Russian version of the BAT (Paradis & Zeiber, 1987) was preferred over the long form because it is faster to administer, and still addresses the most important domains of linguistic functioning. Tasks within the short form are classified into five categories based on the domain of language abilities assessed:

1. Auditory comprehension (pointing, simple and semi-complex commands, verbal auditory discrimination, syntactic comprehension, listening comprehension).
2. Reading (reading aloud, reading comprehension for words, reading comprehension for sentences, reading comprehension for paragraphs).
3. Repetition (repetition of words, repetition of sentences, series).
4. Naming (naming, sentence construction, semantic opposites).
5. Metalinguistic ability (synonyms, antonyms, lexical decision).

As recommended by Paradis (1987), the following tasks from the original test were not included: complex commands, verbal fluency, and description of pictures, mental arithmetic, writing, derivational morphology, morphological opposites, grammaticality judgements, semantic acceptability, and semantic categories.

The adapted version of the test administered in this study did not include tasks on spontaneous speech, copying, and dictation. These tasks were excluded from the short form of the test because they were not pertinent to the design of the larger studies from which the BAT data were obtained (Hallowell et al., 2006; Heuer et al., 2006; Ivanova et al., 2007).

The following changes were made to the test items and/or scoring in the adapted version of the BAT:

1. Verbal auditory discrimination – item #59. Instead of the word "detka" (child) the word "kletka" (cage) was given as the target because, in modern Russian, the word "rebenok" is used to denote child, rather than the original word "detka" (frequency counts 557.01 and 4.71 ipm, respectively, taken from Sharoff, 2002, 2005).
2. Verbal auditory discrimination – item #62. Both the picture of a snake (head of a snake with an open mouth) and the picture of a bottle of poison were considered as correct responses to the target word "yad" (poison).
3. Semantic opposites – item #314. Both words "nepravilnuu" (incorrect) and "lojnuu" (wrong) were considered correct answers (opposites) to the word "pravilnuu" (correct) (as in the antonyms task, the word "nepravilnuu" [incorrect] was actually given as an antonym to the word "pravilnuu" [correct]).
4. Semantic opposites – item #321. The words "malenkiu" (small) and "nizkiu" (short), which have almost the same meaning in Russian, were both considered as correct opposites to the word "vusokiu" (tall).
5. Semantic opposites – item #323. The words "hudou" (slim) and "tonkiu" (thin), which have almost the same meaning in Russian, were both considered to be correct opposites for the word "tolstuu" (stout).
6. Reading – items #387–392 (reading text). A spelling error "otpravi- lis" was changed to "otpravilis" and a grammatical mistake was corrected ("sobrat ovoscheu" was changed to "sobirat ovoschi"); the text was retyped using a font and size similar to the original.
7. Reading comprehension for words – item #415. Instead of the word "shauka" (bathtub) the word "mauka" (shirt) was given as the target. The word "shauka" was not presented because it is the only homophone among target words in this task and its second meaning (gang of thieves) is more common.

Modifications to some items were based on a detailed analysis of the Russian version of the test by the two investigators and discussion of the problematic items with three Russian colleagues. Any modifications in scoring were made following data collection for the first 40 participants, based on observations made during testing and because a disproportionate number of incorrect responses had been recorded for some items. After the scoring procedure was modified to accommodate irregularities in performance the patients' responses were re-evaluated. The same version of the short form of the BAT in Russian was administered to all patients. Administration was consistent across participants, following the published instructions (Paradis & Zeiber, 1987).

Procedure

The adapted version of the short form of the BAT in Russian was administered to each participant. Except for modifications described in detail above, the test items were presented and scored exactly as described by Paradis (1987) for the BAT in general and in the Russian test booklet (Paradis & Zeiber, 1987).

The first author and another neuropsychologist administered the test. The examiners tested the first three patients together to clarify any ambiguities in the administration procedure and recording of responses. All further administration was performed independently. The first author scored all tests after administration.

RESULTS

On average the short form of the BAT took 60 to 90 minutes to administer. Most of the patients completed the test in two sessions. It was observed during assessment that some of the items within the verbal auditory discrimination and reading comprehension for words tasks included ambiguous images. Visually confusing stimuli led to erroneous responses by some patients, even when the linguistic item was comprehended correctly. For instance, when participants were instructed to point to a picture of a "harbour" in item #52, six patients (7%) could not visually distinguish a picture of a harbour from a picture of a castle; eight (10%) stated that the picture of a harbour was not present in the multiple-choice visual array although their verbal interpretation of the target word "harbour" reflected accurate comprehension of its meaning. Similar problems were encountered for items #41 (picture of "bangs" [a fringe hairstyle] looks more like a picture of a face), #59 and #417 (pictures of a "cage" and a "net" look similar), #409 (the picture of a storm that is supposed to correspond to the target word "rolling motion" is too visually complex and it is difficult to interpret what is happening in the picture). The scoring scheme for these items was not altered, since the decision on whether the error was due to linguistic deficits or difficulties in visually recognising an image was based on subjective post-hoc judgements.

Description of patients' performance

Scores (percentage correct) across categories of tasks for patients with different types of aphasia and various levels of severity of language impairment are summarised in Table 1. Post-hoc analysis revealed that there was no significant difference in the overall scores between performance of stroke and TBI patients, $F(1, 80) = .522$, $MSE = .038$, $p = .47$, $\omega^2 = .01$. There were also no differences in performance across tasks between the two groups.

TABLE 1
Mean percentage of items correct for different types of aphasia and severity of language impairment across task categories

		Auditory Compre-hension		Reading		Naming		Repetition		Meta-linguistic		Total score	
	N	Mean %	*SD* %	Mean %	*SD* %	Mean %	*SD* %	Mean %	*SD* %	Mean %	*SD* %	Mean %	*SD* %
Type of aphasia													
Anterior	44	82	14	71.2	28	65.7	33	68.1	33	79.7	25	72.9	22
Broca's	32	81.5	16	71.1	30	65.3	35	64.8	35	78.8	26	73.9	24
Transcortical motor	12	83.3	11	71.6	26	66.9	29	76.9	26	84	12	77.7	17
Posterior	39	86.4	10	82	18	76.1	28	80	24	88.3	18	81.6	14
Wernicke's	9	76	10	65.9	24	53.3	29	56.4	29	79.7	11	68.1	16
Transcortical sensory	25	89.4	8	86.8	12	83.3	22	89.3	9	93.8	7	88.7	8
Conduction	3	87.9	10	82.6	14	70.5	46	55.8	44	70	43	76.1	25
Anomic	2	93.1	1	93.9	0	97.1	4	97.5	4	95	0	94.8	1
Severity													
Mild	22	94.7	4	92.8	5	94.3	13	93.3	8	96.1	4	93.9	4
Moderate	34	87.2	7	85	12	86.6	18	82.5	17	89.9	9	85.7	9
Severe	21	74.2	11	61.3	21	68.6	26	56.6	29	76.6	18	65.5	14
Very severe	6	61.3	10	18.8	16	21	11	8.8	14	35.8	31	32.2	8
Overall	83	84.1	13	76.3	24	70.6	30	73.5	29	84.3	20	78.9	19

Reliability

Internal consistency of the BAT was evaluated using the Kuder-Richardson correlation for dichotomous items (analogous to the Cronbach's alpha used for multiple choice items; Fishman & Galguera, 2003). This correlation coefficient may be interpreted as the mean of all possible split-half coefficients and therefore is preferred over those based on the selection of specific items considered to constitute each half of a test. Internal consistency was examined within each task category and within each task (Table 2). Descriptive statistics (mean, percent correct, standard deviation, and range) for each task and category, along with standard error of measurement, are also provided in Table 2.

Validity

Across all five task categories there was a significant difference between groups of patients with different levels of severity of language impairment, $F(3, 79) = 73.57$, $MSE = .01$, $p < .001$, $\omega^2 = .72$. Post-hoc analysis revealed that overall BAT scores decreased linearly with subjectively rated severity of language impairment, linear-trend contrast, $F(1, 79) = 207.44$, $MSE = .01$, $p < .001$. It was not possible to investigate differences in performance between specific subtypes of aphasia due to insufficient representation within some subtypes. Patients with posterior aphasia

TABLE 2
Descriptive statistics, internal-consistency measures (Kuder-Richardson correlation), and Standard Error of Measurement for the tasks and task categories of the BAT

	Maximum score possible	Mean	Percent correct	*SD*	Range	Kuder-Richardson correlation	Standard Error of Measurement*
Auditory Comprehension	80	67.24	84.09	10.09	32–80	0.922	2.82
Pointing	10	9.8	98	0.68	6–10	0.649	0.40
Simple and semi-complex commands	10	8.66	86.6	1.51	5–10	0.685	0.85
Verbal Auditory Discrimination	18	16.93	94.06	1.28	13–18	0.444	0.95
Syntactic Comprehension	37	28.55	77.16	6.9	0–10	0.915	2.01
Listening Comprehension	5	3.33	66.6	1.41	0–5	0.708	0.76
Reading	46	35.1	76.3	11.25	0–46	0.929	3.00
Reading	20	15.34	76.7	4.54	0–20	0.926	1.24
Reading Comprehension for Words	10	8.54	85.4	1.93	0–10	0.499	1.37
Reading Comprehension for Sentences	10	6.83	68.3	2.92	0–10	0.74	1.49
Reading Comprehension for Paragraph	6	4.37	72.83	1.94	0–6	0.796	0.88
TOTAL RECEPTIVE SCORE	126	102.3	81.19	20.26	47–126	0.956	4.25
Repetition	40	29.39	73.48	11.73	0–40	0.96	2.35
Repetition of Words	30	22.64	75.46	8.86	0–30	0.955	1.88
Repetition of Sentences	7	4.28	61.14	2.73	0–7	0.911	0.81
Series	3	2.47	82.33	1.04	0–3	0.876	0.37
Naming	35	24.71	70.6	10.77	0–35	0.965	2.01
Naming	20	15.8	79	6.31	0–20	0.963	1.21
Sentence Construction	5	2.24	44.8	1.78	0–5	0.78	0.83
Semantic Opposites	10	6.67	66.7	3.62	0–10	0.92	1.02
TOTAL EXPRESSIVE SCORE	75	54.1	72.13	21.3	0–75	0.976	3.30
Metalinguistic Ability	40	35.18	87.95	4.8	22–40	0.808	2.10
Synonyms	5	4.08	81.6	1.4	0–5	0.846	0.55
Antonyms	5	3.46	69.2	1.82	0–5	0.853	0.70
Lexical Decision	30	27.47	91.56	2.68	18–30	0.637	1.61
TOTAL SCORE	241	190.1	78.88	46.25	50–239	0.974	7.46

*The following formula was used to calculate the Standard Error of Measurement: $SD*(1 - r_{xx})^{1/2}$, where *SD* is the standard deviation of all the scores in the sample, and r_{xx} is the reliability coefficient for the test.

subtypes performed slightly better (averaging an overall mean of 82%) than patients with anterior aphasia subtypes (72%), $F(1, 81) = 4.35$, $MSE = .036$ $p = .04$, $\omega^2 = .04$.

Item analysis

The mean score obtained on an item across all participants can be regarded as an indication of its level of difficulty, with higher scores indicating easier items. Ideally, the distribution of item difficulty should be close to normal and the range should be from 0 to approximately 90 to 100% (most likely corresponding to normal performance). This ensures that the test includes items with varying degrees of difficulty (assuming that other sources of variation are eliminated through analyses of item consistency, as discussed below). Consequently, this helps ensure the test is sensitive to small differences in aphasia severity.

The frequency distribution of items according to their difficulty level is presented in Figure 1. The mean item difficulty was .8 (with standard deviation of .14). A total of 103 items (43%) were answered correctly by at least 90% of patients; of those items, 23 (9.5%) had a mean score of 1, meaning that they were answered correctly by every patient in the sample. These were mostly items from tasks on pointing, simple and semi-complex commands, auditory comprehension, reading comprehension for words, and lexical decision. The following six items (2.5%) were answered correctly by less than 50% of participants: sentence construction #309 (mean score .28), #299 (.35), #304 (.45), listening comprehension #363 (.32), reading comprehension of sentences #423 (.42), and repetition of sentences #256 (.47).

Item consistency (or item reliability) was examined by re-computing the instrument's reliability *n* times, deleting a different item from the instrument each time (alpha-if-item-deleted) (Fishman & Galguera, 2003). When deleting a particular

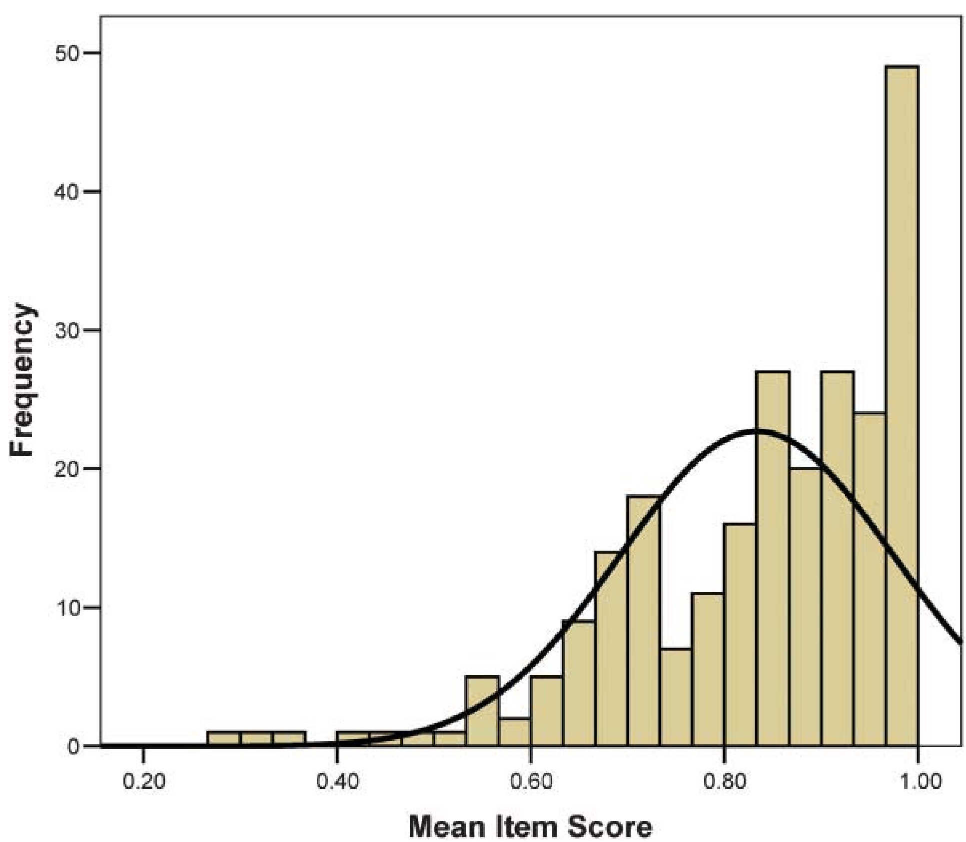

Figure 1. Distribution of items according to mean score. Mean item score of 1 means that those items were answered correctly by all patients (100%) and a score of .5 means that 50% of patients responded correctly.

item causes the instrument's overall reliability to rise, that item can be considered to be a poor item, as the instrument becomes more consistent without it. On the other hand, if deletion of an item causes the overall instrument reliability to fall, then the item is desirable, since it contributes to the instrument's reliability.

The following items were found to decrease the overall reliability of the particular task in measuring the selected construct (split-half reliability coefficient increased when the item was deleted from the set): pointing #25, 28; simple and semi-complex commands #37; verbal auditory discrimination #51, 58, 62, 64, 65; synonyms #158; antonyms #163; repetition #227, 231; lexical decision #206, 232, 238; reading comprehension for words #416. Items that had a mean score of 1 and thus no variance could not be included in this analysis.

Item discriminability was explored using an item validity estimation procedure. It is computed by correlating scores on a particular item to the total score minus the score on that item (corrected item total correlation). A bell-shaped curve with a mean in the 0.60–0.80 is desired for this statistic (Fishman & Galguera, 2003). A moderate to high correlation with the overall score reflects that the item measures the same underlying construct as that appraised by the instrument overall. Special attention should be paid to items that demonstrate a high correlation (around .9) with the overall score, as that might be an indication of the redundancy of that particular item. As mentioned previously, it is desirable for items to measure a range of behaviours from the domain being assessed.

Item discriminability as measured by the corrected item total correlation for each task and task category are presented in Table 3. Items that have a negative or zero correlation with the total score (reflecting a lack of a relationship with the overall task score) are also listed in Table 3.

DISCUSSION

Results from administration of the modified short form of the BAT to a sample of 83 Russian-speaking patients with aphasia were analysed. Overall, the test differentiated patients according to level of severity of language impairment. All five task categories and most tasks within these categories had high internal consistency values: above .7 for syntactic comprehension, listening comprehension, reading, reading comprehension for sentences, reading comprehension for paragraph, repetition of words, repetition of sentences, series, naming, sentence construction, semantic opposites, synonyms, and antonyms, as measured by the Kuder-Richardson correlation. These findings bode well for the test's usefulness in quantifying language impairments in patients with aphasia. However, it would be appropriate to address key problems with the current short form of the BAT in Russian in future research efforts.

Reliability was a concern (internal consistency below .7, as measured by Kuder-Richardson correlation) in the following tasks: pointing, simple and semi-complex commands, verbal auditory discrimination, lexical decision, and reading comprehension for words. A primary reason why reliability was not high may be that a perfect score was obtained on several items in these tasks. In other words, a number of items in these tasks were not sensitive in detecting aphasic language deficits, while others were sensitive, which led to low split half consistency.

Detailed item analysis supported this conclusion. In general, participants performed near ceiling levels. Mean item difficulty was just .8 and 43% of items

TABLE 3
Item discriminability: Corrected item total correlation for each task and task category

	Mean corrected item total correlation	*SD* of the corrected item total correlation	Range of the corrected item total correlation	Items with negative or zero corrected item-total correlation*
Auditory Comprehension	0.32	0.21	−0.09 to 0.65	–
Pointing	0.21	0.31	−0.03 to 0.79	# 23, 27, 28, 29, 31, 32
Simple and semi-complex commands	0.26	0.24	0–0.59	# 33, 34, 35, 36
Verbal Auditory Discrimination	0.13	0.13	−0.11 to 0.4	# 51, 58, 60, 62, 63, 64, 65
Syntactic Comprehension	0.47	0.10	0.19–0.61	
Listening Comprehension	0.48	0.10	0.37–0.6	
Reading	0.44	0.20	−0.07 to 0.71	–
Reading	0.61	0.10	0.41–0.76	
Reading Comprehension for Words	0.18	0.19	−0.09 to 0.51	# 410, 413, 416
Reading Comprehension for Sentences	0.41	0.12	0.14–0.57	
Reading Comprehension for Paragraph	0.55	0.09	0.39–0.66	
TOTAL RECEPTIVE SCORE	0.35	0.20	−0.09 to 0.64	–
Repetition	0.63	0.14	0.23–0.87	–
Repetition of Words	0.67	0.15	0.3–0.89	
Repetition of Sentences	0.74	0.06	0.63–0.81	
Series	0.77	0.05	0.72–0.81	
Naming	0.66	0.11	0.36–0.82	–
Naming	0.75	0.08	0.6–0.87	
Sentence Construction	0.56	0.10	0.42–0.69	
Semantic Opposites	0.70	0.08	0.57–0.83	
TOTAL EXPRESSIVE SCORE	0.61	0.12	0.13–0.83	–
Metalinguistic Ability	0.23	0.24	−0.1 to 0.7	–
Synonyms	0.66	0.13	0.51–0.81	
Antonyms	0.67	0.10	0.52–0.81	
Lexical Decision	0.15	0.17	−0.06 to 0.55	# 194, 196, 198, 206, 212, 214, 216, 224, 226, 230, 238, 242, 244, 246, 252.
TOTAL SCORE	*0.34*	*0.22*	*−0.15 to 0.79*	–

*Items that have a corrected item total correlation less than .1 are listed.

had a mean difficulty equal or above .9; the average score obtained by participants with aphasia was 80%. Based on aphasia diagnosis and severity ratings, the observed distribution of item difficulty and the average score of 80% indicated a lack of desired sensitivity in characterising the deficits of patients with moderate to mild aphasia, who comprised a large proportion (65%) of the sample in this study. Also, tasks that had low internal consistency tended to include many items with poor discriminability (negative or zero corrected item total correlation). This suggests that these items were not related to the overall task score.

At present it is not possible to establish criterion validity of the test directly, as no data have been collected on control participants without neurological impairment. It is not possible to evaluate concurrent validity based on other assessment batteries, as there are no standardised language tests available in Russian. Test–retest reliability was not examined, as the test has not been administered twice to the same participants under the conditions where no change in underlying language abilities could be assumed. Inter-rater reliability also has not been addressed.

DIRECTIONS FOR FUTURE RESEARCH

The current study represents an important first step in initiating development of standardised, valid, and reliable language assessment batteries in Russian. Some tasks in the modified short form of the Russian BAT require further revision and validation. Additionally, it is important to collect data on the spontaneous speech, copying, and dictation tasks not included in this study and to evaluate their psychometric properties.

Those tasks with both poor reliability and low mean corrected item total correlation (pointing, simple and semi-complex commands, verbal auditory discrimination, lexical decision, and reading comprehension for words) require further development. The reliability of these tasks could be improved by removing items that all patients performed correctly, and subsequently including more difficult items. In some instances, it is advisable to substitute relatively easy tasks with analogous but more difficult ones from the long form of the test. For example, instead of simple and semi-complex commands, the complex commands task could be used. It is not possible to discern whether low discriminability and consistency of some of the items is due to unbalanced difficulty within or across subtests or to other validity concerns mentioned above. Therefore, the test's psychometric properties should be further studied after the modifications to task difficulty are made.

Given that the internal consistency of most tasks was high, the longer tasks (especially syntactic comprehension, repetition of words, and lexical decision) could be shortened by reducing the number of items in each. While shortening some of the existing tasks, other tasks from the long form of the test (e.g., spontaneous speech and writing) could be added to this modified short form to make the assessment tool more comprehensive. The overall duration of the test could also be shortened through implementation of ceiling/floor rules for the longer tasks.

Revision or exclusion of some of the visual stimuli from the verbal discrimination and reading comprehension for words tasks would be appropriate given that some patients did not recognise the pictures used in these tasks (items # 41, 52, 59, 409, and 417). In general, it is important to control for visual characteristics of images used to assess linguistic comprehension in aphasia (Heuer & Hallowell, 2007).

Apart from the modifications described above, to standardise the test and establish criterion validity, normative data must be collected on healthy individuals without any history of neurological or language impairment. Also, aside from criterion validity, other indices should be considered when examining the instrument's ability to distinguish normal from impaired performance. Even though the mean scores of the two groups might be significantly different, there is likely to be overlap between them (Ross & Wertz, 2003, 2004). When making a clinical decision based on test results it is important to consider the degree of overlap (proportion of patients with aphasia scoring at or above the minimum expected score for healthy controls) and/or the index of determination (the degree to which being diagnosed with aphasia compared to having normal language skills predicts performance on a test) (Ross & Wertz, 2003). Psychometric properties of a test such as sensitivity (percent of patients with aphasia who perform below a cut-off score for normal performance) and specificity (the proportion of patients without aphasia/healthy individuals who obtain results above the cut-off for normal language abilities) are also valuable when evaluating how much test scores increase the post-test probability of a correct diagnosis (Maxwell & Satake, 2006; Ross & Wertz, 2004). Ideally these psychometric properties should be established and reported. To provide comprehensive evaluation of the test's psychometric properties, inter-rater and test–retest reliability must be investigated as well. Furthermore, careful control for participant knowledge of more than one language would be important for future applications of the test with bilingual patients, as knowledge of additional languages may impact performance on some language tasks (Roberts, 1998).

In summary, a preliminary evidence base for clinicians and investigators using the Russian version of the BAT has been established and will serve as a foundation for future studies of the psychometric properties of the BAT in Russian (and perhaps other languages). Recommendations for modification and scoring have been made with the aim of further developing standardised aphasia language tests in Russian.[3] Efforts to develop standardised testing will enhance clinical procedures and augment the scientific merit of research involving speakers of Russian with aphasia.

Manuscript received 12 June 2007
Manuscript accepted 9 November 2007
First published online 5 February 2008

REFERENCES

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Prospect Heights, IL: Waveland Press.

Cvetkova, L. S., Axytina, T. V., & Pulaeva, N. M. (1981). *Kolichestvennaya ocenka rechi y bolnux s aphasieu* [Quantitative language assessment in patients with aphasia]. Moscow: Izdatelstvo Moskovskogo Gosydarstvennogo Yniversiteta.

Fishman, J. A., & Galguera, T. (2003). *Introduction to test construction in social and behavioral sciences*. Oxford, UK: Rowman & Littlefield Publishers.

Hallowell, B. (in press). Strategic design of protocols to evaluate vision in research on aphasia and related disorders. *Aphasiology*.

[3] The authors of this article do not intend to co-author or publish any modified version of the BAT.

Hallowell, B., & Chapey, R. (in press). Introduction to language intervention strategies in adult aphasia. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related communication disorders* (5th ed.). Philadelphia: Lippincott Williams & Wilkins.

Hallowell, B., Kruse, H., Shklovsky, V. M., Ivanova, M. V., & Emeliyanova, M. A. (2006, November). *Validity of eye tracking assessment of auditory comprehension in aphasia*. Poster presented at the American Speech-Language-Hearing Association Convention. Miami Beach, FL.

Heuer, S., & Hallowell, B. (2007). An evaluation of multiple-choice test images for comprehension assessment in aphasia. *Aphasiology*, *21*, 883–900.

Heuer, S., Hallowell, B., Ivanova, M. V., Kruse, H., Shklovsky, V. M., & Emeliyanova, M. A. (2006, November). *Multiple-choice stimulus properties influence verbal/nonverbal task performance in aphasia*. Paper presented at the American Speech-Language-Hearing Association Convention. Miami Beach, FL.

Hyvärinen, L., Näsänen, R., & Laurinen, P. (1980). New visual acuity test for pre-school children. *Acta Ophthalmologica*, *58*, 507–511.

Ivanova, M. V., Hallowell, B., Kruse, H., Shklovsky, V. M., & Emeliyanova, M. A. (2007, November). *Indexing changes in auditory comprehension in aphasia using eye tracking*. Poster presented at American Speech-Language-Hearing Association Convention. Boston, MA.

Luria, A. R. (1976). *Osnovu neuropsichologii* [Fundamentals of neuropsychology]. Moscow: Izdatelstvo Moskovskogo Gosydarstvennogo Yniversiteta.

Maxwell, D. L., & Satake, E. (2006). *Research and statistical methods in communication sciences and disorders*. Boston: Thomson/Delmar Learning.

Paradis, M. (1987). *The assessment of bilingual aphasia*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Paradis, M., & Zeiber, T. (1987). *Bilingual Aphasia Test (Russian version)*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Roberts, P. M. (1998). Clinical research needs and issues in bilingual aphasia. *Aphasiology*, *12*, 119–130.

Roberts, P. M. (2001). Aphasia assessment and treatment in bilingual and multicultural populations. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related communication disorders* (4th ed., pp. 208–232). Philadelphia: Lippincott Williams & Wilkins.

Ross, K. B., & Wertz, R. T. (2003). Discriminative validity of selected measures for differentiating normal from aphasic performance. *American Journal of Speech-Language Pathology*, *12*, 312–319.

Ross, K. B., & Wertz, R. T. (2004). Accuracy of formal tests for diagnosing mild aphasia: An application of evidence-based medicine. *Aphasiology*, *18*, 337–355.

Sharoff, S. (2002). *The frequency dictionary for Russian*. Retrieved 10 September 2007, from http://www.comp.leeds.ac.uk/ssharoff/frqlist/lemma.al.zip.

Sharoff, S. (2005). Methods and tools for development of the Russian Reference Corpus. In D. Archer, A. Wilson, & P. Rayson (Eds.), *Corpus linguistics around the world* (pp. 167–180). Amsterdam: Rodopi.

Shklovsky, V. M. (2003). Koncepciya neuroreabilitacii bolnux c posledstviyami insulta [Conception of neurorehabilitation in post-stroke patients]. *Jyrnal nevrologii i psixiatrii im. S.S. Korsakova*, *9*, 13–37.

Spreen, O., & Risser, A. H. (2003). *Assessment of aphasia*. Oxford, UK: Oxford University Press.