DOI: 10.4274/Tnd.16779

Original Article / *Özgün Araştırma*

Validity, Reliability and Standardization Study of the Language Assessment Test for Aphasia

Afazi Dil Değerlendirme Testi'nin Geçerlik, Güvenirlik ve Standardizasyon Çalışması

Bülent Toğram, İlknur Maviş

Anadolu University Center for Speech and Language Pathology, Eskişehir, Turkey

Summary

Objective: Aphasia assessment is the first step towards a well-founded language therapy. Language tests need to consider cultural as well as typological linguistic aspects of a given language. This study was designed to determine the standardization, validity and reliability of Language Assessment Test for Aphasia, which consists of eight subtests including spontaneous speech and language, auditory comprehension, repetition, naming, reading, grammar, speech acts, and writing.

Material and Method: The test was administered to 282 healthy participants and 92 aphasic participants in age, education and gender-matched groups. The validity study of the test was investigated with analysis of content, structure and criterion-related validity. For reliability of the test, the analysis of internal consistency, stability and equivalence reliability was conducted. The influence of variables on healhty participants' sub-test scores, test score and language score were examined. According to significant differences, norms and cut-off scores based on language score were determined.

Results: The group with aphasia performed significantly lower than healthy participants on subtest, test and language scores. The test scores of the healthy group were mostly affected by age and educational level but not by gender. According to significant differences, age and educational level for both groups were determined based on . Considering age and educational levels, the reference values for the cut-off scores were presented.

Discussion: The test was found to be a highly reliable and valid aphasia test for Turkish- speaking aphasic patients either in Turkey or other Turkish communities around the world. *(Turkish Journal of Neurology 2012; 18:96-103)*

Key Words: Aphasia, language assessment test for aphasia, validity, reliability and standardization

Özet

Amaç: Afazi değerlendirmesi iyi yapılandırılmış müdahale programının ilk basamağıdır. Dil testlerinin, uygulandığı toplumun kültürel özellikleri kadar tipolojik olarak dilsel özelliklerini de gözetmesi gerekmektedir. Türk toplumunun kültürel ve dile özgü tipolojik özelliklerine uygun geliştirilen Afazi Dil Değerlendirme Testi, afazili bireyleri dil puanı açısından sağlıklılardan anlamlı olarak ayırt etmeyi hedeflemektedir. Bu çalışmanın amacı, spontane dil ve konuşma, işitsel anlama, tekrarlama, adlandırma, okuma, söz eylemler, dilbilgisi ve yazma olmak üzere sekiz alt testten oluşan testin, geçerlik güvenirlik ve standardizasyon çalışmasını gerçekleştirmektir.

Gereç ve Yöntem: Test, yaş, eğitim düzeyi ve cinsiyete göre gruplanmış 282 sağlıklı ve 92 afazili katılımcıya uygulanmıştır. Testin geçerliği içerik, yapı ve ölçüt geçerliği analizleri, güvenirliği ise iç tutarlılık, istikrarlılık ve eşdeğerlik analizleri ile gerçekleştirilmiştir. Sağlıklı katılımcıların alt test, test ve dil puanları üzerinde değişkenlerin etkisi incelenmiş ve anlamlı farklılığa göre norm ve kesme değer puanları belirlenmiştir.

Bulgular: Afazili grup, alt test puanları, test puanları ve dil puanlarında sağlıklı katılımcılardan anlamlı olarak daha düşük performans sergilemiştir. Sağlıklı grubun test puanlarının yaş ve eğitim düzeyinden etkilendiği, ancak cinsiyetin etkisinin olmadığı belirlenmiştir. Anlamlı farklılığa göre ortaya çıkan yaş ve eğitim düzeyine göre testin norm ve kesme değer puanları belirlenmiştir.

Sonuç: Afazi Dil Değerlendirme Testi'nin Türkiye'de ve dünyada Türkçe konuşan afazili hastalar için yüksek derecede güvenilir ve geçerli bir afazi testi olduğu doğrulanmıştır. *(Türk Nöroloji Dergisi 2012; 18:96-103)*

Anahtar Kelimeler: Afazi, Afazi Dil Değerlendirme testi (ADD), geçerlik, güvenirlik ve standardizasyon

Address for Correspondence/Yazışma Adresi: Bülent Toğram MD, Anadolu University Center for Speech and Language Pathology, Eskişehir, Turkey
Gsm.: +90 542 593 96 28 E-mail: btogram@anadolu.edu.tr
Received/Geliş Tarihi: 23.07.2012 **Accepted/Kabul Tarihi:** 29.08.2012

Introduction

Aphasia is the most destructive impairment among all that stroke causes. Twenty-one to 38 percent of all acute stage stroke patients are found to have aphasia (1). Aphasia is defined as a loss or impairment in the process of understanding or generating language symbols following acquired brain damage widely affecting cortical and subcortical networks of the linguistically dominant hemisphere (2,3). Language and communication disorders in aphasia differ based on the demographic and socio-cultural factors, as well as the location, severity and size of the lesion causing the brain damage, and need to be evaluated carefully. Aphasia not only has an effect on the individual's personality and social life, but influences relations with family and others, therefore quality of life. A multidisciplinary team including a doctor, a physiotherapist, an occupational therapist and a language and speech therapist is needed to improve the quality of life of the aphasic individual and help them re-adjusting to their social environment. During rehabilitation, the language and speech therapist should help improve the patient's communication skills and substantially decrease the impact of the condition. Assessment of aphasia is based on the detailed evaluation of the patient's language skills in all areas and is the first step in a well-structured intervention program that depends heavily on the individual's requirements (4). It is well known that when assessing aphasia, areas of deficiency should be determined by appropriate tests and observation results will help in predicting prognosis and helping plan therapeutic objectives. Naturally, the tool used in assessment should measure all language skills, allow comparison with a norm group and have a standard application.

Most of the aphasia tests used throughout the world (Boston Diagnostic Aphasia Examination- BDAE) (5); (Western Aphasia Battery- WAB) (6); (English Aachen Aphasia Test- EAAT) (7) were developed in North America. Some of these tests were translated into other languages, but using diagnostic tests developed in other countries may create problems interpreting results due to cultural, demographic and linguistic differences. dir. Experts who are aware of this difficulty feel the need to develop a test that is appropriate to their own language and cultural structure (8, 9). Language tests must take into consideration the cultural characteristics as well as the typologically linguistic features of the society they are being applied in. Turkish, a Ural-Altaic language, is differentiated from the Indo-European language family. Therefore, translations of the aphasia assessment tests in other languages are not suitable for Turkish. Although there are very few validated aphasia screening tests in Turkish [Frenchay Aphasia Screening Test (10); Gülhane Aphasia Test–2 (GAT-2) (11); Ege Aphasia Test EAT (12)], there is no comprehensive aphasia assessment tool that evaluates in detail the language skills that are presumed to be impaired in aphasia and which has been standardized, validated and tested for reliability.

Although bedside evaluation usually identify the type of aphasia, the formal language evaluation of a language and speech therapist is important for identifying the fine details of the deficiency, planning of the appropriate therapy and describing the improvement potential of the patient. Therefore, each language element must be tested individually and in detail. The overall objective of this study is to test the validity, reliability and standardization of the Aphasia Language Assessment Test (ALA), developed as a specific language test for Turkish language and culture for the abovementioned requirements. Hence, the answers to the following questions were seeked: (a) What is the validity of ALA? (b) What is the reliability of ALA? (c) What are the subtest scores, test scores and language scores of the participant groups (healthy and aphasic)? (d) What is the correlation of the subtest scores, test scores and language scores with the age and education level of the healthy participants group? (e) What are the norm scores and cut-off score of ALA?

Material and Method

Participants in the Study

A total of 374 participants (282 healthy and 92 aphasic) were enrolled in the study. Healthy subjects lived in Eskişehir and environs (Afyon, Bursa, Bilecik, Kütahya, Ankara and Düzce), whereas aphasic subjects lived in Eskişehir and various cities in Turkey (İstanbul, Kars, Ankara, İzmir, Konya, Şanlıurfa, Sakarya, Kocaeli, Bursa, İzmir).

The distribution of age and education level was q u i t e similar in both the healthy and aphasic groups, as well as the distribution rates of female / male subjects. Mean time after stroke for aphasic participants was found to be approximately 2 years.

Data Collection Tool

In this study data were collected using "Aphasia Language Assessment Test (ALA)" (13), developed from 2005, to provide information for language and speech therapists specializing in diagnosing aphasia, a language disorder due to left brain damage.

Criteria for Participants

Participants for this study were identified utilizing criterion dependent sample selection. Healthy participants had presented at the Neurology and Physical Therapy Outpatient Clinic at a medical school for complaints other than brain damage, and were included in the study after having been identified as having 'no neurological problems' by specialists. Inclusion criteria included no history of stroke or brain damage, no progressive central nervous system disorder (Alzheimer's, Parkinsonizm etc), no psychiatric disorders,

sensorial problems, drug addiction, thyroid disorders and B12 deficiency, not taking any drugs known to influence cognition, mother tongue to be Turkish and willing to take part in the study. Aphasic participants, on the other hand, had to be examined by a neurologist and the damage causing aphasia reported to be focal and localized in the left brain and due to stroke, diagnosed by a specialized language and speech therapist as 'aphasia', and able to take the test; aphasic participants also had to have no progressive central nervous system disorders (Alzheimer's, Parkinsonizm etc) before or after the stroke, no sensorial problems accompanying aphasia, no history of language / speech / learning problem prior to the stroke, mother tongue to be Turkish, and currently not receiving any therapy at the date of assessment.

Aphasia Language Assessment Test (ALA)

ALA aims to identify the language area-related performance in individuals suffering from left brain damage, to diagnose aphasia and to help select appropriate therapeutical targets. ALA consists of 8 subtests including assessment of spontaneous language and speech (SLS), auditory understanding (AU), repetition (Rp), naming (N), reading (R), grammar (G), speech act (SA) and writing (W). Three distinct reactions are defined to score ALA: Correct (C) / Independent Reaction (2 Points), Missing / Insufficient / Assisted Reaction (M) (1 Point) and Incorrect (I) Reaction or No Response (NR) (0 Point). The test was administered to healthy participants in a quiet and comfortable room with no negative environmental factors, in the neurology clinic of a university, between 20 and 30 minutes. On the other hand, the test was administered to aphasic participants in the language and speech disorders center of a university, in 60 to 90 minutes. Two types of scores, including test score (TSCORE) and language score (LSCORE), were calculated in this study. Test score consists of the sum of all the subtests of ALA (292 points), and language test consists of the sum of the subtests 'spontaneous language and speech assessment, auditory understanding assessment, repetition assessment and naming assessment' (162 points).

Procedure

Study for validity, reliability and standardization of ALA

The study for content validity for ALA included developing the test and obtaining expert opinion, pilot study; investigating the differences in ALA subtest scores and test scores of participant groups in the study and differences between the scores of end groups, correlations between test scores and subtest scores and factor analysis in the study for structural validity; and investigating the correlation of ALA with GAT-2 and EAT in the study for criteria validity (14, 15). The correlation of ALA with both tests was investigated separately with 30 aphasic cases each. Correlations were found between the common subtests of ALA and EAT, auditory understanding, repetition, naming, reading and writing and common subtests of ALA and GAT-2, language-cognition assessment / awareness, auditory understanding, automatic speech, repetition and naming.

The reliability study for ALA involved calculation of the mean of item-total score correlation coefficients for each subtest composing the test and Cronbach alpha coefficients for each subtest and test in general, to determine internal reliability, a test-retest method was applied to determine consistency, and inter-rater reliability coefficient was calculated to determine equivalence. Thirty healthy and 7 aphasic subjects were included in the test-retest reliability study. ALA was readministered to the healthy participants one week after the initial test, whereas the aphasic participants were readministered ALA one to two weeks after the initial assessment. Video images of randomly selected 30 subjects among aphasic participants were scored by a specialized language and speech therapist for the inter-rater realiability study. Spearmann correlation coefficient was calculated to estimate test – retest and inter-rater reliability.

Standardized tests are also named formal tests, as administering and scoring involve standard rules (16). For any test to be considered norm referenced and for psychometric adequacy, the specified population must be investigated in terms of at least three of the age, gender, education and socioeconomic level characteristics (17). Therefore, ALA was administered to 282 healthy participants fulfilling the

Table 1. Demographic information of healthy and aphasic participants

		Healthy (n=282)	Aphasic (n=92)
Gender	Female	145	38
	Male	137	54
Age (years)	23-44	88	17
	45-59	102	32
	60-74	73	31
	75+	19	12
		(51.9±13.7)	(57.3±14.1)
Education (years)	NL	19	15
	1-5	114	29
	1-8	38	18
	1-11	76	17
	12+	35	13
		(7.9±4.1)	(7.2±4.6)
Weeks after stroke	0-1	-	13
(=29.2± 50.1)	1-4	-	19
	4-12	-	14
	12+	-	46
Type of aphasia	Fluent	-	26
	Non-fluent	-	46
	Global	-	20

NL: Not literate

Table 2. Difference in test scores of participant groups

Subtests	group	n	$\overline{X}$	SS	SE	SD	t	p
SLS	Healthy	282	30.98	1.68	.10	372	27.87	.000*
	Aphasic	92	15.88	8.63	.89			
AU	Healthy	282	64.12	2.94	.17	372	17.97	.000*
	Aphasic	92	40.60	21.40	2.23			
Rp	Healthy	282	19.70	.88	.05	372	27.25	.000*
	Aphasic	92	7.35	7.46	.77			
N	Healthy	282	43.84	.79	.047	372	35.83	.000*
	Aphasic	92	12.32	14.74	1.53			
R	Healthy	282	45.28	12.29	.73	372	20.16	.000*
	Aphasic	92	13.69	15.15	1.57			
G	Healthy	282	19.59	1.20	.071	372	35.37	.000*
	Aphasic	92	4.77	6.73	.70			
WA	Healthy	282	19.86	.63	.03	372	36.65	.000*
	Aphasic	92	4.89	6.78	.70			
W	Healthy	282	37.40	9.35	.55	372	19.91	.000*
	Aphasic	92	11.80	14.07	1.46			
TSCORE	Healthy	282	280.80	26.31	1.56	372	31.73	.000*
	Aphasic	92	111.33	77.10	8.03			
LSCORE	Healthy	282	158.65	5.04	.30	372	31.35	.000*
	Aphasic	92	76.17	43.39	4.52			

SLS: Spontaneous language and speech; AU: Auditory understanding; Rp: Repetition; N: Naming; R: Reading; G: Grammar; WA: Word act; W: Writing; TSCORE: Test score; LSCORE: Language score *p<0.001

required criteria (norm group), data obtained were evaluated in terms of the subjects' ages, education and gender and differences of variables were studied. Test performances of healthy and aphasic participants were compared based on variables and groups with significant differences, cut-off value scores of healthy participants were calculated based on test and language scored and norm scores were determined. Moreover, in this study, cut-off value score are given based on the language score found to be appropriate to identify aphasia.

Discussion

Findings on the Validity of ALA

ADD was developed on the basis of "Diagnostic Assessment Criteria for Aphasia", recommended by Hedge (18).

Expert opinion on the form and content of the test was obtained from six academicians and one language and speech therapist practicing in this field. The high Kappa Agreement Coefficient between the experts (99) demonstrated that the test items adequately represented the features they aimed to measure.

Table 3. Eigen Values and Variance Percentages of ALA (Aphasia Language Asessment) Factors

Factor	Eigen Value	Percentage of Variance	Accumulated Percentage of Variance
1	22.374	77.2	77.2
2	2.227	7.7	84.8
3	1.720	5.9	90.8

The pilot study for ALA was conducted with 25 (14 female / 11 male) healthy subjects without any neurological problems. Data from the study show that the participants' mean 'test scores' were quite high (mean=282.72±7.3), suggesting, as predicted, that healthy participants gave correct answers to test items without much difficulty.

Differences between the subtest, test scores and language scores of healthy and aphasic participant groups to determine the structural validity of ALA, and a significant difference was found (p<0.001) (Table 2). This suggested that ALA could distinguish between healthy and aphasic participants, identify aphasia and had the appropriate structure to measure its objective.

Another finding in favor of structural validity was the significant difference in TSCORE between the top 27% (n=76) and bottom 27% (n=76) tier of the healthy participants [t(150)=-.897, p>.001].

Pearson's correlation coefficients were analyzed to find the relationship between the ALA subtest scores and test scores of healthy participants, and substantial positive correlation was found between substest scores for 'spontaneous language and speech (r=.76, p<.01), auditory understanding (r=.77, p<.01), repetition (r=.50, p<.01), naming (r=.31, p<.01), reading (r=.99, p<.01), grammar (r=.56, p<.01), word acts (r=.28, p<.01) and writing (r=.97, p<.01)' and test scores of healthy participants. A finding of high positive correlation also indicates that the structural validity of ALA is high.

Subtest scores for ALA were analyzed with Principal Component Factor Analysis, and subtests with a factor load above .60 and loading components were identified. Table 3 shows Eigen values and variance percentages of the factors found.

Published literature stresses that for a measuring tool to show that it measures a single conceptual structure, it has to describe at least 40% of the total variance of the first factor and the weight of other factors has to decrease progressively (19). In this study, Factor 1 includes 'spontaneous language and speech (.69-.73), repetition (.82), naming (.78-.86), grammar (.80) and word act assessment (.81)', Factor 2 includes reading (.76-.84) and writing (.79-.83) assessment, whereas Factor 3 includes auditory understanding (.74-.81) assessment. Findings show that ALA is three dimensional and all factors en masse measure a single conceptual structure.

The correlation between common subtests for ALA and EAT, auditory understanding, repetition, naming, reading and writing was investigated for the analysis of criterion validity and the correlation between the tests (.90, .89, .98, .90 and .94, respectively) was found to be significant (p<.01) and high (15); on the other hand, correlation between common subtests for ALA and GAT-2, language – cognition assessment / awareness, auditory understanding, automatic speech, repetition and naming was found to be high (.94, .76, .90, .94, .94, respectively), as well (14). The high correlation of ALA with the common subtests for both tests provides strong evidence for criterion validity.

Findings on the Reliability of ALA

ALA was analyzed within three reliability categories: internal consistency, stability and equivalence.

Total scores for the subtest items of ALA and coefficients of correlation change between .36 and .88 (=.56) for language and speech, between .40 and .67 (=.51) for auditory understanding, between .56 and .96 (=.76) for repetition, between .73 and .93 (=.82) for naming, between .63 and .83 (=.74) for reading, between .76 and .85 (=.80) for grammar, between .80 and .91 (=.85) for word act, and between .78 and .95 (=.86) for writing. Hence, the internal consistency for all subtests of ALA is found to be high. Cronbach alpha coefficients estimated for the ALA subtests (Spontaneous language and speech .94, Auditory understanding .97, Repetition .97, Naming .99, Reading .99 Grammar .97, Word act .98, Writing .99 and ALA .99) show that the internal consistency of ALA is fairly well.

It was also shown that there was a substantial consistency in the test – retest reliability of ALA (r=0.88, p<.001) and the inter-rater reliability of the test was considerably high (r=0.97, p<.001), as well.

Findings on the Standardization of ALA

Pearson's correlation coefficient was calculated to discover the effect of age and education on test performance in the healthy participants group (Table 4), and the subtest, test and language scores of healthy participants were found to increase with education level, and have an inverse relationship with age.

Findings for norm scores and cut-off value scores of ALA

One-Way Analysis of Variance was performed to determine the difference between subtest, test scores and language scores of healthy participants, based on education status and age, and in case of a difference, Tukey's HSD Post Hoc Test was performed to determine which groups it arises from. For education status, a difference between the not literate group and the 1-5 years of education group and any other group in terms of spontaneous language and speech, auditory understanding, reading, language score and test score is significant. Subtests where the not literate group differentiates from the other groups were repetition [F(4)=13.3, p<.0001], naming [F(4)=2.72, p<.03], grammar [F(4)=13.3, p<.0001], word act [F(4)= 9.3, p<.0001] and writing [F(4)=721, p<.0001]. It was seen that there was no difference between the 23-44 years and 45-59 years age groups for subtests of ALA, in terms of age, and all other age groups had differences from

Table 4. Correlation of subtest scores, test scores and language scores of healthy participants with age and education level

		Age	Education
SLS	r	-.58**	.56**
AU	r	-.49**	.50**
Rp	r	-.47**	.28**
N	r	-.20**	.13*
R	r	-.48**	.49**
G	r	-.28**	.27**
WA	r	-.16**	.18**
W	r	-.38**	.44**
TSCORE	r	-.49**	.51**
LSCORE	r	-,59**	,55**

SLS: Spontaneous language and speech; AU: Auditory understanding; Rp: Repetition; N: Naming; R: Reading; G: Grammar; WA: Word act; W: Writing; TSCORE: Test score; LSCORE: Language score Correlation is significant at *0.05; ** 0.01 level

the others. It was noted that only the subtests of naming and word act had differences in the 75 years and over age group from the 23-44 years and 45-59 years age groups.

Results of the t-test analysis measuring the effect of the variable of gender based on the scores of the healthy participants group, test scores [t(280)=-.381, p>.05] and language scores [t(280)=-.853, p>.05] did not show a significant difference for the variable of gender.

Norm and cut-off Value scores for ALA

Norm scores for ALA were determined using the study variables age and education. The groups that had a statistically significant difference in language scores based on the variables age and education, were divided into three (23-59 years, 60-74 years and 75+ years), (NL; 1-5 years and 6+ years) (Table 5), mean language scores of healthy participants were calculated for variables and norm scores were generated.

Receiver Operating Characteristic (ROC) curve was used to determine the cut-off value scores to distinguish between healthy and aphasic participants and to diagnose aphasia. The age and education level groups based on language score and used in determining norm scores were also used in determining cut-off value scores and results are shown in Tables 6 and 7.

Based on the cut-off value scores, for the language score of ALA, 23-59 year old individuals with an education of 1-5

Table 5. Norm scores of healthy participants based on language scores

Age group	Education Level (NL)		(1-5 yr)		(6+ yr)	
	$\overline{X}$	SS	$\overline{X}$	SS	$\overline{X}$	SS
(23-59 yr)	-	-	159.2	3.04	160.9	1.70
(60-74 yr)	146.8	4.5	157.8	2.8	160.13	2.3
(75+ yr)	143.9	8.6	153.2	3.5	-	-

NL: not literate

Table 6. Area under the curve (AUC) and standard error (SE) values based on language scores

Age group	NL AUC (SE)	Education Level 1-5 yr AUC (SE)	6+ yr AUC (SE)
23-59 yr	-	.999 (.004)	.999 (.002)
60-74 yr	.958 (.042)	.979(.015)	.969 (.045)
75+	1.000 (.000)	1.000 (.000)	-

NL: not literate

Table 7. Cut-off value scores (CVS), positive predictive value (PPV), negative predictive value (NPV) based on language scores

Age group	Education Level NL	1-5 yr	6+ yr
23-59 yr	-	CVS: 152	CVS: 155
		%95 CI:0.942-1.000	%95 CI:0.976-1.000
		Se:1.00	Se:0.97
		Sp:0.98	Sp: 1.00
		PPV:0.94	PPV: 1.00
		NPV:1.00	NPV:0.99
60-74 yr	CVS: 127	CVS: 145	CVS: 152
	%95 CI:0.838-0.995	%95 CI:0.920-0.997	%95 CI:0.792-0.990
	Se:0.92	Se:0.93	Se:0.93
	Sp:1.00	Sp:1.00	Sp:1.00
	PPV: 1.00	PPV:1.00	PPV:1.00
	NPV: 0.86	NPV:0.96	NPV:0.89
75+ yr	CVS: 63	CVS: 118	-
	%95 CI:0.734-1.000	%95 CI:0.780-1.000	
	Se:1.00	Se:1.00	
	Sp:1.00	Sp:1.00	
	PPV: 1.00	PPV:1.00	
	NPV: 1.00	NPV:1.00	

NL: not literate; Sn: Sensitivity; Sp: Specificity; CI: Confidence interval

years scoring 152 and below, 23-59 year old individuals with an education of 6 years or above and scoring 155 and below, 60-74 year old individuals who are not literate and who score 127 and under, 60-74 year old individuals with 1-5 years of education and a score of 145 and below, 60-74 year old individuals with 6 or more years of education and a score of 152 and below, individuals in the 75+ years age group who are not literate and who score 63 and below, and individuals in the 75+ years age group with 1-5 years of education and a score of 118 and below can be diagnosed with "aphasia due to left brain damage".

Discussion

When data from participant groups are analyzed in the evaluation of ALA, scores of healthy participants are found to be high, as expected, whereas scores of aphasic participants are found to be clearly lower compared to the healthy participants. In particular, mean language score predicted to be affected in aphasia resulting from left brain damage was found to be 76 in aphasic participants and 159, very close to the highest possible score (162) in healthy participants. This clear difference reveals that ALA distinguishes aphasic participants from healthy participants. When test score data were reviewed, mean score for the healthy participants group (281) was found to be much higher than the mean score for the aphasic group (111). Hence, the high value of standard deviation for the aphasic participants group in both language score (SD: 43.40) and test score (SD: 33.96) shows the variability of the participants' scores in ALA. The reason for this variability is thought to be the participation of various types of aphasia (particularly severe non-fluent aphasia) in the study; however, when aphasic and healthy participants are compared for the means of language scores and test scores, the difference in language score was found to be 82 (approximately 16 SD), and the difference in test score to be 15 (approximately 1/2 SD). Language scores of aphasic participants are expected to be much lower than healthy participants because language score consists of the sum of the subtests spontaneous language and speech, auditory understanding, repetition and naming, which are expected to be affected by a possible language problem following left brain damage. On the other hand, the reason of the high test score standard deviation in healthy participants (SD: 26.31) is thought to be the not literate group participating in the study, because 90 points of the highest possible score of 282 in ALA consists of the subtests reading and writing assessment. In light of all these findings, ALA was found to be suitable to identify the strong and weak points of the individual in various language areas.

Participants in this study reflect the social and cultural reality in Turkey in terms of age, education and gender. The number of subjects who were not literate and who had a lower education level was higher among older participants compared to other age groups. Due to these characteristics, the potential impact of demographic variables on test scores is higher. Based on our findings, the test performance of healthy participants increase with the education level of participants, as expected ($r=.13$: $p<.05$ to $.56$: $p<.01$); on the other hand, scores got higher as age got lower ($r=-.16$:$p<.05$ to $-.59$: $p<.01$). Literature search showed that the findings of this study was confirmed by a study conducted with healthy participants where test scores for the Western Aphasia Battery (WAB) decreased as age increased ($r=-.63$, $p<0.01$) (20), and in the standardization study of the Frenchay Aphasia Screening Test with healthy participants, where scores for the aphasia test decreased as age increased (10). In a study of the Boston Aphasia Diagnostic Examination, it was revealed that while education had a significant impact on most skills, age had an impact on some skills (21). In the norm study conducted with the Spanish Boston Aphasia Diagnostic Examination, education was found to have a significant effect on most subtests, and the only significant difference among age groups existed among the subtests word-picture matching and serial writing (22). In a study for the adaptation of the Boston Aphasia Diagnostic Examination for the Brazilian population, the variable of age was found to have a significant effect on some subtests (complex thought, naming a picture, understanding a spelled word, naming by writing and writing a dictated sentence) and level of education was found to have a significant effect on all subtests except naming by response, recognizing words and word-picture matching (23). In the norm study for the Korean version of the Western Aphasia Battery, findings of age and education having an effect on test scores (24), and in the standardization study of the English version of the Aachen Aphasia Test (7) the significant association between age and most subtests seem to verify the findings of the present study. However, studies inconsistent with the findings of this study were found in literature; for example, in a study conducted with the Boston Aphasia Diagnostic Examination, Whitworth and Larson (25) report that they have not found an important difference in their sample for the impact of education, and Miller et al. (7) express that there was no significant correlation between the scores and education in the standardization study for the English version of the Aachen Aphasia Test.

Gender was not found to have an effect on the performance of healthy participants in this study. Findings of numerous studies in literature report that there is no difference for the variable of gender (25, 22, 24, 10, 26, 7). When these findings are taken into consideration, this study suggests that language tests, as a diagnostic tool, should reflect age and education factors.

Norm scores of ALA are determined for groups with significant differences in language scores for age and education

level variables of healthy participants. Highest scores of both norm scores and cut-off value scores of ALA were found to belong to the youngest age group and those participants with the highest education level, whereas the lowest scores belonged to the oldest group and those participants who were not literate. It should be remembered that patients with scores just above the cut-off value may have mild aphasia, and the individual's personal complaints in terms of communication, language and speech problems should be considered. In a literature review, in the standardization, validity and reliability studies for aphasia tests, it was seen that cut-off values were determined using one or both of the age and education variables of participants; moreoever, when cut-off values were being determined, 1 or 2 standard deviation (SD) below mean was used or ROC analysis method was used similar to this study. Literature shows that in a study conducted with the original Boston Aphasia Diagnostic Examination, cut-off value was determined as 2 SD below mean (27). In the adaptation study of the Boston Aphasia Diagnostic Examination for the Brazilian population (23), cut-off value scores were suggested for only education level, even though there were differences between age groups; in the norm study for the Korean version of the Western Aphasia Battery (24), it was reported that two and three groups were created for statistically significance in age and education level, respectively, and cut-off value scores were estimated for these groups. In the study aiming to identify the German normative data of the Aphasia Checklist battery, Kalbe et al. (26) were reported to develop cut-off value scores on 1 SD below mean of healthy participants.

This study has some limitations, including participants being from Eskişehir and environs (Bursa, Afyon, Bilecik, Kütahya), younger healthy participants having a higher education level than the older healthy participants, not literate participants being included in the mean test score calculations of healthy participants, and number of female, not literate, healthy participants being slightly higher than the corresponding male participants.

References

1. Berthier ML. Poststroke aphasia: epidemiology, pathophysiology and treatment. Drugs Aging 2005;22:163-82.
2. Tyrer JH, Jordan JM. Rationale and short description of the Queensland University Aphasia Test(Q.U.A.T.). Int J Lang Commun Disord 1971;6:164-72.
3. McNeil MR, Pratt SR. Defining aphasia:some theoretical and clinical implications of operating from a formal definition. Aphasiology 2001;15:900-11.
4. Spreen O, Risser AH. Assessment of aphasia. New York: Oxford University Press, 2003.
5. Goodglass H, Kaplan E. The Assessment of aphasia and related disorders. Philadelphia: Lea & Febiger, 1972.
6. Shewan CM, Kertesz A. Reliability and validity characteristics of the Western Aphasia Battery (WAB). J Speech Hear Disord 1980;45:308-24.
7. Miller N, Willmes K, DeBleser R. The psychometric properties of the English language version of the Aachen Aphasia Test (EAAT). Aphasiology 2000;14:683-722.
8. Helm-Estabrooks N, Albert LM. Manual aphasia and aphasia therapy. Texas: Pro-Ed, 1991.
9. Helm-Estabrooks N, Albert LM. Manual aphasia and aphasia therapy. 2nd ed. Texas:Pro-Ed, 2004.
10. Göçer-March E. Frenchay Afazi Tarama Testi: Türk nöroloji hastaları için bir standardizasyon çalışması. Türk Psikol Derg 1996;11:56-63.
11. Tanrıdağ O, Maviş İ, Topbaş S. GAT-2:Gülhane Afazi Testi-2. Ankara: Detay Yayıncılık; 2011.
12. Atamaz F, Yagiz On A, Durmaz B. Ege Aphasia Test: Description of the test and performance in normal subjects. Türk Fiz Tıp Rehab Derg. 2007;53:5-10.
13. Maviş İ, Toğram B. Afazi Dil Değerlendirme Testi (ADD) kullanım yönergesi. Ankara: Detay Yayınları, 2009.
14. Duru H. İnmeli bireylerin afazi dil değerlendirme testi ve gülhane afazi testi-2 performanslarının ilişkisi. Eskişehir: Anadolu Üniversitesi Sağlık Bilimleri Fakültesi Yayımlanmamış yüksek lisans projesi; 2012.
15. Çıkan G. İnmeli bireylerin afazi dil değerlendirme testi ve ege afazi testi performanslarının ilişkisi. Eskişehir: Anadolu Üniversitesi Sağlık Bilimleri Fakültesi Yayımlanmamış yüksek lisans projesi; 2012.
16. Shipley KG, McAfee JG. Assessment in speech-language pathology a resource manual. 3rd Ed. New York: Thomson-Delmar Learning, 2004.
17. American Psychological Association (APA). Standards for educational and psychological testing. Washington: Author, 1999.
18. Hedge MN. Pocket guide to assessment in speech-language pathology. 2nd ed. San Diego: Singular Thompson Learning Inc, 2001.
19. Şencan H. Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlik. Ankara:Seçkin Yayıncılık, 2005.
20. Ross KB, Wertz RT. Possible demographic influences on differentiating normal from aphasic performance. J Commun Disord 2001;34:115-30.
21. Rosselli M, Ardila A, Florez A, Castro C. Normative data on the Boston Diagnostic Aphasia Examination in a Spanish-speaking population. J Clin Exp Neuropsychol 1990;12:313-22.
22. Pineda DA, Roselli M, Ardila M, Mejia SE, Romero MG, Perez C. The Boston Diagnostic Aphasia Examination- Spanish version: The influence of demographic variables. J Int Neuropsychol Soc 2000;6:802-14.
23. Radanovic M, Mansur LL, Scaff M. Normative data for the Brazilian population in the Boston Diagnostic Aphasia Examination: Influence of schooling. Braz J Med Biol Res 2004;37:1731-38.
24. Kim H, Na DL. Normative data on Korean Version of the Western Aphasia Battery. J Clin Exp Neuropsychol 2004;26:1011-20.
25. Whitworth RH, Larson CM. Differential diagnosis and staging of Alzheimer's disease with an aphasia battery. Neuropsychiatry Neuropsychol Behavioral Neurol 1989;1:255-65.
26. Kalbe E, Reinhold N, Brand M, Markowitsch HJ, Kessler J. A new test battery to assess aphasic disturbances and associated cognitive dysfunctions- German normative data on the Aphasia Check List. J Clin Exp Neuropsychol 2005;27:779-94.
27. Borod JC, Goodglass H, Kaplan E. Normative data on the Boston Diagnostic Aphasia Examination, Parietal Lobe Battery, and the Boston Naming Test. J Clin Neurospychol 1980;2:209-15.