

From informal to formal: the preliminary psychometric evaluation of the short aphasia test for Gulf Arabic speakers (SATG)

Article

Accepted Version

Altaib, M. K., Falouda, M. and Meteyard, L. (2020) From informal to formal: the preliminary psychometric evaluation of the short aphasia test for Gulf Arabic speakers (SATG). *Aphasiology*. ISSN 1464-5041 doi: <https://doi.org/10.1080/02687038.2020.1765303> Available at <http://centaur.reading.ac.uk/91327/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1080/02687038.2020.1765303>

Publisher: Taylor and Francis

including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Altaib, M.K., Falouda, M. & Meteyard, L. (2020) From informal to formal: the preliminary psychometric evaluation of the short aphasia test for Gulf Arabic speaker (SATG). *Aphasiology*, published online May 27th 2020; <https://doi.org/10.1080/02687038.2020.1765303>

From informal to formal: the preliminary psychometric evaluation of the short aphasia test for Gulf Arabic speaker (SATG)

Authors: Madhawi Khalid Altaib*^{1,3}, Mona Falouda², Lotte Meteyard³

*Corresponding author

¹ *Health Rehabilitation Sciences Department, College of Applied Medical Sciences, King Saudi University, Riyadh, Kingdom of Saudi Arabia, 12371.*

E-mail address: Maltaib@ksu.edu.sa, Twitter Account: @Madhawi_k

² *Communication and Swallowing Disorders Department, Rehabilitation Hospitals, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia, 12231.*

³ *School of Psychology & Clinical Language Sciences, University of Reading, Reading, United Kingdom, RG6 6AL.*

Word count: 6968 (main text and tables) and 7200 (main text, tables and abstract)

Abstract

Background: Speech and language therapists in Gulf Arabic countries still rely on informal aphasia and/or translated Western-language assessments to assess the language proficiency of people with aphasia. However, these tests are not sensitive to the linguistic and cultural features of the Arabic language, which may lead to inaccurate diagnosis. This paper describes the preliminary development and the preliminary psychometric evaluation of the short aphasia test for Gulf Arabic speaker (SATG).

Method: The aim was to develop and preliminary assess the psychometric properties of the SATG. Three phases determined whether subtests and tasks were culturally and linguistically appropriate for Gulf Arabic populations. The test consists of six sections that assess different language skills: semi-spontaneous speech, auditory comprehension, repetition, naming, automatic speech, recitation, reading and writing. Together, these aim to detect the absence or presence of aphasia and provide a broad classification of aphasia syndrome (fluent and non-fluent).

Result: The SATG takes 20 minutes to complete. It was administered to 37 healthy adult controls and 31 people with aphasia post-stroke. In this pilot study the SATG demonstrated good to excellent reliability over time and from one clinician to another. The SATG was found to have face, content and concurrent validity.

Conclusion: Preliminary results indicate that the SATG is a reliable and valid aphasia assessment. Further study is needed to examine the efficacy of the SATG to screen for the presence of aphasia (i.e. differentiate between those with and without aphasia post-stroke), distinguish severity levels for aphasia, and to improve standardisation with a wider range of control participants.

Key words:

Assessment, aphasia, Arabic, Saudi Arabia, psychometric, test

Acknowledgements:

Research Center of the Female Scientific and Medical Colleges, Deanship of Scientific Research, King Saud University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

Aphasia is the loss or impairment of the language system following brain damage, and the most common cause of aphasia is stroke, a cardiovascular accident (CVA) (Benson & Ardila, 1996). Approximately one-third of people who suffer a stroke develop aphasia (Brady, Kelly, Godwin, Enderby, & Campbell, 2016; Engelter et al., 2006). In order to establish the presence of aphasia, speech and language therapists (SLTs) perform a series of tests that target the patient's language abilities. Those tests usually consist of multiple subtests that include spontaneous/conversational speech, auditory comprehension, repetition, object naming, reading and writing. Comprehensive batteries of these tests for American English can be found in the Boston Diagnostic Aphasia Evaluation, BDAE, (H. Goodglass & Kaplan, 1972) and the Western Aphasia Battery, WAB, (Kertesz, 1982) and for British English, the Comprehensive Aphasia Test, CAT, (Swinburn, Porter, & Howard, 2004). The history of aphasia testing began in the nineteenth century, when clinicians were mostly dependent on informal assessment and observation to assess people with aphasia (PWA) (Byng, Kay, Edmundson, & Scott, 1990; Howard, Swinburn, & Porter, 2009). The first reported aphasia test battery was developed by Rieger in 1888 (Byng et al., 1990). Following that, many types of assessments were developed. Comprehensive aphasia batteries focus on classifying aphasia types (e.g. the WAB, the BDAE) or providing a profile of impairment against a population of individuals with aphasia (e.g. the CAT). Other tests focus on assessing functional communication ability such as the communicative abilities in daily living test, CADL, (Holland, 1980). Additionally, there are tests that assess a specific area of language ability, such as the Boston naming test, BNT, (Kaplan, Goodglass, & Weintraub, 2001). There are screening assessments that are designed to provide a quick evaluation of language impairment, usually performed at the hospital bedside, e.g. Frenchay Aphasia Screening Test, FAST, (Enderby, Wood, & Wade, 2006). Recently, another format has been introduced which aims to fill the gap between lengthy test batteries and quick aphasia screening tests, for example the Quick aphasia battery, QAB, (Wilson, Eriksson, Schneck, & Lucanie, 2018). (Enderby, Wood, & Wade, 2006)

Most existing aphasia assessment tests are biased towards Western cultures and are available in English or European languages. They were developed for English-speaking populations, and materials were designed in relation to Western culture, such as the CAT, WAB and BDAE (H. Goodglass & Kaplan, 1972; Kertesz, 1982; Swinburn et al., 2004). In recent years, due to increasing awareness, there has been an increasing demand to develop assessments for PWA in languages other than English, including Arabic. In Gulf Arabic

countries, speech and language therapy is a developing field with a lack of assessment materials for PWA. As yet, there are no gold standard or fully developed, standardised Aphasia assessment batteries for Gulf Arabic. Saudi SLTs are still reliant on informal assessments and/or translated Western-language assessments which are not sensitive to the linguistic and cultural features of the Arabic language. This may lead to inaccurate diagnosis (Khamis-Dakwar & Froud, 2012). The only study that investigated the assessment material used by SLTs was conducted by (Khoja, 2017), who used a survey to identify the formal and informal assessment procedure used by SLTs in Saudi Arabia. She stated that most SLTs working with PWA reported the use of Western formal assessments such as the Boston Diagnostic Aphasia Evaluation, BDAE (Harold Goodglass, Kaplan, & Barresi, 2001) and Western Aphasia Battery, WAB (Kertesz, 1982) without cross-cultural adaptation or standardisation. They also reported the use of informal assessment either by itself or in addition to non-adapted Western formal assessment (Khoja, 2017). Direct translation of Western aphasia assessment is problematic since the Western norms are no longer usable, essentially making the direct translation an informal assessment. This issue could be overcome by establishing new test norms for the translation. However, this would entail new issues such as linguistic differences between the Arabic language and the native language of the translated test, including differences in phonology, morphology and syntax that may no longer then represent the original manipulations, e.g. for length or complexity of the stimuli (Ivanova & Hallowell, 2013).

The Arabic language is listed as the fourth most spoken language globally, with more than 280 million people speaking Arabic as their first language (Saad & Ashour, 2010). In 2011, Beveridge and Bak reported that only 5 aphasia related articles were published in relation to the Arabic-speaking population. Few attempts have been made to either develop aphasia assessment or cross culturally adopt an existing Western test. In Jordan for example, there is the Bilingual aphasia test, BAT, (Paradis & El Halees, 1989) in Arabic-Jordanian/Levantine dialect. Likewise, in Lebanon, there is the development of the bedside version of the Arabic diagnostic aphasia battery (A-DAB-1) in Arabic-Lebanese/Levantine dialect (Al-Thalaya et al., 2017). Also, in Egypt there are two assessments: the Kasr El-Aini Arabic Aphasia Test, KAAT, (Hassanein et al., 2002), and there is progress in developing a cross culturally adapted version of CAT for Arabic-Egyptian dialect speakers (Abou El-Ella et al., 2013). However, due to the diversity of Arabic dialects, using non-Gulf Arabic dialects such as Egyptian or Levantine dialect for Gulf Arabic speaking populations is problematic (Beveridge & Bak, 2011). In Gulf countries, there is the adaptation of the object and action

naming battery into Saudi/Gulf Arabic dialect (Alyahya & Druks, 2015) and there is progress in the development of the comprehensive Aphasia battery for Qatari/Gulf Arabic speaking populations (Khwaileh, Mustafawi, Howard, & Herbert, 2016).

In this paper, we report the initial development and pilot of the Short Aphasia test for Gulf Arabic speaking populations (SATG), informed by the current practice of SLTs in Saudi Arabia and by existing aphasia batteries. The aim was to develop a short formal assessment for Aphasia, and make a preliminary evaluation of its reliability and validity.

Methods

Study design:

Participants

Control group

A convenience sample of healthy adults were recruited to act as a control group in order to inform test development, validate the selection of subtests and items for subtests, and to provide pilot data for standardisation (i.e. data to use as a comparison for scores from PWA). The inclusion criteria of health adult were, all Gulf-Arabic speaker health adult aged from 21 to (80+) years old, with no history of stroke or any speech and language disorder

A total of 37 (12 males and 25 females) healthy individuals, ranging in age from 21 to 71 years (mean: 38.40 years; standard deviation: 13.23), were recruited as a control group. Due to the nature of this project (time and resources) we could not match to our PWA group for age, gender and education level. All participants were native Gulf Arabic speakers. Please refer to Table (1) for the demographic information of the control group

Table 1 The demographic information of control group

Gender		Male	Female	Total
(Number)		12	25	37
Age means		47.58 years	34 years	38.40 years
(Standard deviation)		(14.48)	(10.20)	(12.23)
Nationality		Saudi Arabia (12)	Kuwait (7)	Bahrain (5)
		UAE (6)	Qatar (3)	Oman (4)
Education level	High school	1	2	3
	Bachelor	7	14	21

(Number)	Master	4	9	13
-----------------	---------------	---	---	----

PWA group

The inclusion criteria for participants with Aphasia (PWA) were: Gulf Arabic speakers, post stroke, presenting with a language disorder as observed/noted by a member of the Multi-Disciplinary Team and with a referral for the Speech and Language Therapy clinic. Due to the absence of a gold-standard aphasia assessment for Gulf-Arabic, we have taken the approach of validating our test against relevant other characteristics (Rutjes et al, 2007). Both the referral reason and the report of the principle Speech and Language Therapist (SLTs) were used as evidence for a diagnosis of Aphasia. Both clinicians involved in the study data collection were blind to the SLT report and had no involvement in the initial diagnosis of PWA.

A convenience sample of 31 (16 males and 15 females) PWA were recruited, from four hospitals in Saudi Arabia and one hospital in Kuwait. The group ranged in age from 32 to 68 (mean: 55.83 years; standard deviation: 10.07), 29 with Left CVA stroke and 2 with Right CVA stroke. The time since the participants' strokes ranged from 1 to 66 months (mean: 12.25 months; standard deviation: 15.79). Please refer to Table (2) for demographic information of PWA and refer to table (3) for a detailed description of the sample.

Table 2 The demographic information of PWA group

Gender		Male	Female	Total
(Number)		16	15	31
Age mean		57.18 years	54.4 years	55.83 years
(Standard deviation)		(10)	(10.28)	(10.07)
Months post onset mean		16.43 months	7.73 months	12.25 months
(Standard deviation)		(19.14)	(9.98)	(15.79)
Education level	Illiterate	3	8	11
	Intermediate	4	2	6
	High school	7	1	8
	Bachelor	4	2	6
(Number)				

Table 3a: Detailed description of patients' information										
ID	Age (years)	Gender	MPO	Education level	Side of paralysis	Site of lesion (see Table 3b)	Type of Stroke	Handedness	date of testing	Hospital
1	63	M	19	high school	Right	n/a	Lt. MCA stroke	Right	07/11/2016	Saudi
4	45	F	23	illiterate	Right	n/a	Lt. MCA stroke	Right	01/12/2016	Saudi
6	48	M	9	bachelor	Right	n/a	Lt. MCA stroke	Right	07/12/2016	Saudi
8	64	M	24	high school	Right	n/a	Lt. MCA stroke	Right	27/12/2016	Kuwait
9	61	M	23	high school	Right	n/a	Lt. MCA stroke	Right	25/12/2016	Kuwait
11	64	F	11	illiterate	Right	n/a	Lt. MCA stroke	Right	27/12/2016	Kuwait
13	57	M	12	high school	Right	n/a	Lt. MCA stroke	Right	26/12/2016	Kuwait
15	68	M	48	illiterate	n/a	n/a	Lt. MCA stroke	Right	03/01/2017	Saudi
17	60	M	2	illiterate	Right	n/a	Lt. MCA stroke	Right	04/01/2017	Saudi
19	61	M	36	illiterate	Right	n/a	Lt. MCA stroke	Right	12/01/2017	Saudi
21	50	M	12	bachelor	Right	n/a	Lt. MCA stroke	Right	15/01/2017	Saudi
23	34	F	6	bachelor	n/a	n/a	Lt. MCA stroke	Right	17/01/2017	Saudi
25	62	F	4	illiterate	Right	n/a	Lt. MCA stroke	Right	17/01/2017	Saudi
27	40	M	66	bachelor	Right	n/a	Lt. MCA stroke	Right	26/02/2017	Saudi
29	48	F	1	Illiterate	Right	available	Rt. MCA stroke	Right	15/10/2017	Saudi
30	66	F	11	illiterate	Right	available	Lt. MCA stroke	Right	24/10/2017	Saudi
31	66	F	1	illiterate	Left	available	Rt. MCA stroke	Right	25/10/2017	Saudi
32	63	M	3	high school	Right	available	Lt. basal ganglia stroke	Right	04/10/2017	Saudi
33	68	M	2	Intermediate school	Right	available	Lt. MCA stroke	Right	21/11/2017	Saudi
34	61	F	2	Intermediate school	Right	available	Lt. MCA stroke	Right	21/11/2017	Saudi

35	39	F	1	high school	Right	available	Lt. intracerebral haemorrhagic stroke	Right	13/11/2017	Saudi
36	63	M	1	high school	Right	available	Lt. MCA stroke	Right	12/11/2017	Saudi
37	61	F	1	Intermediate school	Right	N/A	Lt. basal ganglia haemorrhagic stroke	Right	21/11/2017	Saudi
38	63	F	12	bachelor	Right	N/A	Lt. basal ganglia stroke	Right	27/11/2017	Saudi
39	49	F	2	Intermediate school	Right	N/A	Lt. MCA stroke	Right	21/11/2017	Saudi
40	32	M	1	high school	Right	available	Lt. MCA stroke	Right	09/01/2018	Saudi
41	56	M	2	Intermediate school	Right	available	Lt. basal ganglia ischemic stroke	Right	10/12/2017	Saudi
42	49	F	4	illiterate	Right	available	Lt. ischemic stroke	Right	07/12/2017	Saudi
43	60	F	1	illiterate	Right	available	Lt. MCA stroke	Right	07+08/01/2018	Saudi
44	49	F	36	Intermediate school	Right	available	Lt. MCA-ACA strokes due to Moya Moya disease	Right	14/01/2018	Saudi
45	61	M	3	bachelor	Right	N/A	Lt. MCA stroke	Right	21+22/01/2018	Saudi

Site of lesion for 13 out of 31 patients:

Table 3b: Detailed description of patients' information (site of lesion)		
ID	Type of Stroke	Site of lesion: (CT scan / MRI study)
29	Rt. MCA stroke	CT scan was done on 05/09/2017 and results revealed: - Acute/subacute ischemic insult in the right MCA territory,

30	Lt. MCA stroke	<p>CT scan was done on 24/11/2016 and results revealed:</p> <ul style="list-style-type: none"> - Near occlusion of left internal carotid artery without full collapse following a tight post-bulbar stenosis due to circumferential soft plaque. No evidence of dissection.
31	Rt. MCA stroke	<p>MRI was done on 23/10/2017 and results revealed:</p> <ul style="list-style-type: none"> - Acute non-haemorrhagic ischemic insults involving the right pons, right inferior parietal lobule and left superior frontal gyrus. Distribution of the ischemic insults is suggestive of embolic phenomena. - Old right frontal operculum infarction with evidence of cortical laminar necrosis and old blood degradation products. - MRA carotids and brain is limited due to lack of IV contrast nevertheless, it demonstrates atherosclerotic changes, moderate to severe right PCA stenosis, otherwise unremarkable.
32	Lt. Basal Ganglia stroke	<p>Study done on 12/07/2017:</p> <ul style="list-style-type: none"> - There is a large intraparenchymal hyperdense area in the left basal ganglia and insular region extending to the left thalamus with surrounding by vasogenic edema , causing mass effect on the left lateral ventricle . Subtle midline shift to the right side has started to develop. - Deep white matter hypodensity seen likely representing small vessels disease however subacute insult cannot be ruled out. - Lacunar infarction in the right thalamus and in the head of left caudate nucleus is noted. - Posterior fossa structures appear unremarkable.
33	Lt. MCA stroke	<p>MRI was done on 28/08/2017 and results revealed:</p> <ul style="list-style-type: none"> - Left MCA territory diffuse hypo-density keeping with acute/subacute infarction. - No acute intracranial haemorrhage. - The ventricles appear grossly unremarkable.

		<ul style="list-style-type: none"> - No mass effect or midline shift. - Posterior fossa structures are grossly unremarkable. - Paranasal sinuses and mastoid air cells are well-aerated.
34	Lt. MCA stroke	<p>done on 19/09/2017 at kingdom hospital:</p> <ul style="list-style-type: none"> - large hypo-density following left MCA territory consistent with extensive left MCA sub-acute territorial infraction, associated mass affect upon the left lateral ventricle and minimal midline shift to right. No haemorrhagic transformation could be seen. No intracranial or intraventricular hematoma. - Right partial hypo-density and encephalomalacia changes suggestive of chronic infraction in the territory of right MCA. - Brainstem and cerebellum showing no focal lesion
35	Lt. intracerebral haemorrhagic stroke	<p>done on 09/10/2017:</p> <ul style="list-style-type: none"> - Left parietal parenchymal haemorrhage dissecting into the ventricular system in association with subarachnoid haemorrhage in the sulci of the left vertex and subdural haemorrhage along the falx and along the right frontal convexity with surrounding edema and right-sided midline shift.
36	Lt. MCA stroke	<p>Study done on 06/09/2017:</p> <ul style="list-style-type: none"> - Extensive left sided acute cerebral infraction affecting whole left middle cerebral artery territory with minimal antero-superior midline shift. - No haemorrhagic transformation noted.
40	Lt. MCA stroke	<p>CT scan was done on 28/12/2017, results revealed:</p> <ul style="list-style-type: none"> - Left parieto-occipital lobe hypo-density likely representing acute/subacute ischemic insult.
41	Lt. basal ganglia ischemic stroke	<p>MRI was done on 16/11/2017 and results revealed:</p>

		<ul style="list-style-type: none"> - Multiple acute/subacute infarcts along the left pons, bilateral occipital poles, right parieto-occipital region and right corona radiata, in keeping with showering emboli from the left ventricular clot. - No frank haemorrhagic transformation.
42	Lt. Ischemic Stroke	<p>CT scan was done on 14/09/2017 and results revealed:</p> <ul style="list-style-type: none"> - there is evidence of hypo-density in the left temporal lobe extending to the peri-trigonal region and to the posterior limb of the ipsilateral internal capsule representing old regional infarct. - A small lacuna is noted at the left basal ganglia related to old lacunar infarct. - Hypo-density is noted in the left upper posterior parietal region in para-midline location representing ischemic insult of undetermined age. - Bilateral periventricular and deep white matter hypo-densities in keeping with chronic microangiopathy. - Acute on top of chronic infarction cannot be ruled out.
43	Lt. MCA stroke	<p>CT scan was done on 14/12/2017 and results revealed:</p> <ul style="list-style-type: none"> - Left frontal lobe hypo-density as described above, in view of patient's presentation and age, findings likely represent subacute/chronic left MCA territorial infarction.
44	Lt. MCA-ACA strokes due to Moya Moya disease	<p>Multiple strokes (left MCA-ACA strokes) due to Moya Moya disease</p> <p>MRI was done on 12/5/ 2015 and results revealed:</p> <ul style="list-style-type: none"> - progress of multiple old infarcts and new acute infarction of the left lateral temporal lobe and left precentral gyrus. - persistent encephalomalacia of bilateral corona radiata and basal ganglia compatible with chronic infarcts

--- Insert Table 3a and 3b here ---

Material

The development process of the SATG:

First phase

Prior to the development of the SATG, SLTs in Saudi Arabia were asked to identify the most frequently used informal aphasia assessment in the clinics. This was reported as “the aphasia diagnostic informal assessment” developed by Saadi Alzahrani in 2003 (Alzahrani, 2003). No norms or standardisation were established for the assessment. Also, to the best of our knowledge, it has never been published or used in research. However, it is very popular among SLTs in Saudi Arabia. Each SLT intended to use the Alzahrani test differently without following the same testing protocol, instructions and/or scoring system.

The Alzahrani test was examined by the first author, including attending assessment sessions in hospitals in Saudi Arabia, Riyadh, while the clinician administered the assessment as it is usually done in clinical practice. There was a detailed examination of each subtest and task and a check of whether it exists in another Western aphasia assessment. Also, Alzahrani was contacted regarding the development framework of his test and he reported that he relied on a combination of multiple assessments to develop the test. After review, the first author (MA) found that the Alzahrani test development was based on two popular comprehensive Western aphasia formal assessment batteries, the WAB and BDAE. Subtests, tasks and items within the subtests were derived from subtests, tasks and items found in WAB and BDAE. Changes from western assessments were made by Alzahrani for cultural and linguistic purposes (for example: the Gregorian calendar was replaced with the Islamic calendar). No picture material was included with the Alzahrani test. The original version developed by Alzahrani included 15 subtests as listed in table 4:

Table 4 The subtests of the original version developed by Alzahrani

The subtests of the original version developed by Alzahrani
1. Spontaneous speech

2. Auditory comprehension
3. Simple and complex commands
4. Yes/No questions
5. Word and sentence repetition
6. Naming
7. Sentence completion
8. Word fluency
9. Responsive naming
10. Automatized sequence and recitation
11. Spoken letter/word to written letter/word matching
12. Spoken words-written word matching
13. Sentence reading comprehension
14. Picture stimulus written word choice matching
15. Writing

Supplementary material 1 provides the original version of the aphasia diagnostic informal assessment developed by Alzahrani; Supplementary material 2 lists subtests of the original version of “the aphasia diagnostic informal assessment” taken from WAB and BDAE with a detailed description of cultural and linguistic modifications.

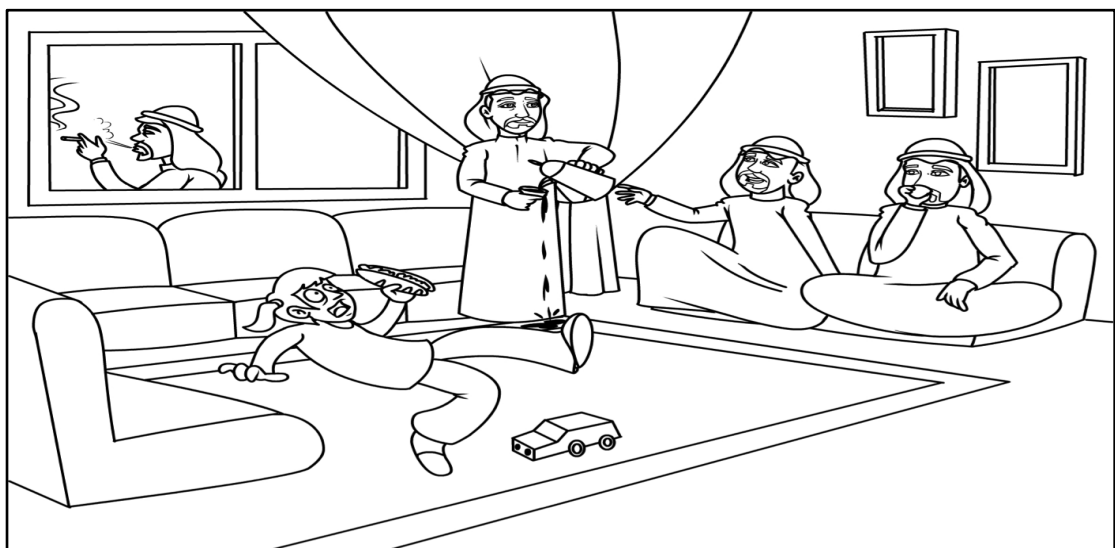
Second phase

In the second phase of the SATG development, the first author met with a group of Saudi SLTs to evaluate the Alzahrani assessment. This evaluation was made considering the objectives of the SATG which include: (1) to develop a short aphasia test for the Gulf Arabic population, (2) the test should be fast and easy to administer (take less than 30 minutes), (3) the test should be suitable for computer based administration, (4) the test should detect the presence or the absence of aphasia, (5) the test should provide an overview of the language profile of PWA, and (6) the test to provide a broad classification of aphasia syndrome (fluent and non-fluent). Thus, it was agreed to write test instructions for subtests that lacked them, for example the Auditory comprehension subtest, and to develop a scoring system that is suitable for a test of short duration. Additionally, there would be modifications, additions and deletions of some subtests from the original Alzahrani test. Changes were suggested for several reasons which include: (1) the assessment was developed in 2003, which makes some items unsuitable; for example, in reading comprehension the subtest item: ‘Saturday is the first day of the week’ should be replaced with ‘Sunday is the first day of the week’, as in

2013 the weekend in Saudi Arabia changed from Thursday and Friday to Friday and Saturday, (2) suitability of the item/stimuli. Items that were judged to be influenced by intelligence or educational level were excluded; for example, in sentence completion the subtest item: ‘The bee absorbs the nectar of...flowers’, (3) cultural and linguistic appropriateness of items; for example, in sentence completion the subtest item: ‘Muslims fast in...Ramadhan’, (4) confusion over some items; for example, in simple and complex commands the subtest item: ‘Put the pen on top of the book and put it back’, (5) redundancy; for example, in yes/no questions the subtest item: ‘Do you live in Hail/not the actual city the patient lives in?’ was removed because another item was targeting the same response. Supplementary material 3 provides the modification of the Alzahrani test and Supplementary material 4 provides the first draft of the SATG.

Furthermore, since there were no pictures within the Alzahrani test, it was decided to include a black and white line drawing picture. A Saudi female artist was employed to draw 48 pictures, with special attention to preparing culturally relevant images. For the picture description task (semi-spontaneous speech section), Saudi SLTs nominated the picture from the Arabic Diagnostic Language test, ADLT, (Asseri, Alsawiti, Alyahya, Alanbar, & Alharbi, n.d.), still under development (Figure 1). This assessment was developed by a group of researchers and clinicians named Areej Asseri, Fadi Alsawiti, Reem Alyahya, Shatha Alanbar and Mohammed Alharbi, in King Fahad Medical City (KFMC). Only the picture for picture description task was directly taken from the KFMC assessment, with scoring criteria designed by the first author.

Figure 1 Picture description (living room gathering) from semi-spontaneous speech section



Third phase

The first draft of the SATG was piloted on 5 healthy normal adults with no history of any neurological disorders aged from 23 to 54 years old, with 3 females and 2 males. The main reasons for piloting were to: (1) determine the testing time, (2) decide on content for the word fluency task (animals or clothing), (3) confirm that items were not influenced by intelligence/education; for example, items within the reading comprehension subtest, and (4) make sure the picture material, instructions and items were familiar and clear. Simultaneously, the SATG was reviewed by three experienced Saudi SLTs (2 years' experience minimum with PWA) for the same reasons and other additional reasons including: (5) checking the cultural and linguistic appropriateness of items and the picture, (6) using their experience with PWA to evaluate subtests and items. After collecting pilot data and the focus group comments, the following changes were made: (1) in simple and complex commands one item was deleted and one item was modified, (2) in repetition three items were deleted, (3) in sentence completion one item were deleted, (4) in picture word naming one item was deleted, and (5) in word fluency the animal category was chosen. Refer to Supplementary material 5 for modifications of the first draft of the SATG, and appendix 1 for the SATG manual.

Fourth phase

The SATG final draft was administered to 37 healthy normal adults. Participants were asked to perform all the tasks within the SATG. The control group response on the semi-spontaneous speech subtest was analysed to establish norms for the MCA scoring system (see Supplementary material 8). The control group responses on word fluency task were averaged to determine norms for word fluency using the animal category. Following this, the SATG was administered to 31 PWA. The order of administration and the instruction was the same for both groups.

Results

Psychometric properties of the SATG

We followed guidelines for the development, psychometric testing and reporting for newly developed assessments in healthcare research (Ivanova & Hallowel, 2013; Simera et al, 2008).

Performance of control group (healthy adults) on SATG:

The data from the control group (healthy adults) were mainly used to allow us to determine the type of responses to expect (i.e. normative data) as this information was not available from existing literature. In particular, the two subtests for spoken production: picture description and word fluency.

The data from the control group were analysed using descriptive statistics. The results show that in 5 out of 14 subtests of the SATG (i.e. responsive naming, automatic speech, letter and word recognition, picture and word matching, and writing on command) the control group reached ceiling performance, for the other subtests the control group performed near ceiling (see Table 5). From Table 5 it can be seen that the mean score of healthy adults is '16.72' and the maximum possible score is '16'. The maximum possible score of '16' was selected to be the scoring criteria as this subtest summed the number of words generated in 1 minute. Due to the lack of the evidence of the amount of words a Gulf-Arabic speaker might generate, we relied on the control group response to decide on the scoring criteria.

Data from the control group did not include illiterate participants, so we were not able to evaluate the influence of this on SATG control scores. It should be noted here that we failed to match the Control group with the PWA group for age and education level. This was due to the convenience sample and resources, as well as the limited time-frame, available to us.

Table 5: Descriptive statistics of the control group performance on SATG 14 subtests:

SATG subtest	maximum possible	mean	SD	minimum	maximum
semi-spontaneous speech	12	8.351	2.1758	2	12
word picture comprehension	16	15.68	0.592	14	16
simple and complex commands	8	7.9	0.296	7	8
Yes and No Questions	11	10.87	0.336	10	11
repetition	12	11.9	0.296	11	12
word naming	21	20.8	0.401	20	21
sentence completion	11	10.78	0.75	10	11
word fluency	16	16.71	4.27	9	27
responsive naming*	8	8	0	8	8
automatic speech*	2	2	0	2	2
letters and word Recognition *	14	14	0	14	14
reading comprehension	7	6.8	0.47	5	7
picture and word matching*	8	8	0	8	8
writing on Commands*	2	2	0	2	2
Total	148	143.378	5.007	130	156

*Reach ceiling

Reliability of the SATG:

The inter-rater and test-retest reliability data was collected from the PWA group to evaluate the reliability of the SATG. For both inter-rater and test-retest reliability 10 out of the 31 PWA were randomly selected. For inter-rater reliability, two trained SLTs were used. One already part of the data collection and the other SLT was randomly selected. Both were blinded to the results. The time interval between ratings was approximately three months. For the test-retest reliability, trained SLTs who were blind to the results were selected, and asked to assess each participant twice with a time interval of 2-8 weeks. None of the participants received therapy in between the two sessions.

Inter-rater reliability:

10 randomly selected participants from the PWA group were rated by two SLTs (one was involved in the study and the second one was not), then both were re-rated after approximately three months. The overall score of the SATG and six sections were computed using the intraclass correlation coefficient (ICC). Criteria for reliability degree were judged using (Fleiss, Levin, & Paik, 2013) (<0.40 = poor reliability; 0.40–0.60 = moderate reliability; 0.61–0.80 = good reliability; 0.81–1.00 = very good reliability). There was very good inter-reliability between the two raters for the PWA on the overall score and across six sections of the SATG. Please refer to Table 6 for the inter-rater reliability of the SATG for PWA group-

Table 6 The inter-rater reliability of the SATG for PWA group

SATG sections	Inter-rater reliability PWA group (10)
Semi-spontaneous speech	0.968
Auditory comprehension	0.992
Repetition	0.986
Naming	0.997
Automated speech	1
Reading and writing	0.999

SATG Overall score	0.998
---------------------------	-------

Test-retest reliability:

10 randomly selected participants from the PWA group were tested twice over a period of 2 to 8 weeks, by the same SLT. The overall score of the SATG and six sections were computed using ICC and Bland and Altman (BA) test of agreement. The result showed very good test-retest reliability for both the PWA and control group on overall score of the SATG and six sections. The BA test of agreement showed narrow Limits of agreement (LOA) for both group control and PWA that is range from (0 to ± 1.91) – that is, variation in scores between the two time points was within two points. Please refer to Table 7 for the test-retest reliability of the SATG for PWA group.

Table 7 The test-retest BA test of Agreement and reliability (ICC) for the PWA group

SATG sections	Bias (Mean difference between two time points)	LOA (Lower to upper)	LOA from zero-point comparison	ICC PWA group (10)
Spontaneous speech	-0.1	-0.71 to 0.51	± 0.61	0.994
Auditory comprehension	-0.3	-2.15 to 1.55	± 1.85	0.995
Repetition	0	0	0	1
Naming	0	-0.923 to 0.923	± 0.923	0.999
Automated speech	-0.1	-0.71 to 0.52	± 0.62	0.949
Reading and writing	0	-0.923 to 0.923	± 0.923	0.999
SATG Overall score	-0.05	-2.4 to 1.4	± 1.91	0.998

Key: LOA = limits of agreement; ICC = intraclass correlation coefficient

Validity of the SATG:

The validity also considers an important section to evaluate the psychometric properties of the SATG. In this preliminary evaluation of the SATG only two types of validities (face and content) validity we were possible due to the lack of an existing, gold-standard assessment (i.e. a reference standard).

Face Validity:

Face validity refers to subjective judgment by experts and/or non-experts to evaluate whether the test matches its purpose and its actual content (Ivanova & Hallowell, 2013). In the SATG,

face validity was obtained by enlisting three SLTs and five caregivers of PWA test takers. They were asked to evaluate whether SATG's content match its purpose or not. Both the group of experts (SLTs) and of non-experts (caregivers) agreed that the SATG targets different modalities within communication skills which mainly become affected in PWA.

Content Validity:

Due to the lack of gold-standard reference aphasia assessment for Gulf-Arabic speakers, the items within the SATG were judged against the BDAE and WAB, to evaluate if each item under each subtest was representative for the subtest .

The SATG assesses word picture comprehension using items from several semantic categories, such as: nouns, actions and colours, similar to both BDAE and WAB. Furthermore, the SATG measures the PWA ability to perform a list of commands with increasing difficulty and to answer yes/no questions.

As in the BDAE and WAB, semi-spontaneous speech was elicited using a picture description task and the picture used within the SATG resembles complex pictures used in other aphasia assessments (i.e. multiple actors and separate events to be described). However, in the SATG, MCA was used to analyse the PWA's production, to allow for easy and fast scoring. Also, the SATG included a conversational observation checklist with simplified fluency and rating scales adapted from the WAB and BDAE.

The repetition section includes word and sentences of increasing length and complexity. In the naming section, the SATG assesses the oral production of PWA using similar tasks to the WAB and BDAE, that is word to picture naming, word fluency, sentence completion and responsive naming.

In reading and writing sections, the SATG also follows the BDAE and WAB by including subtests that assess letter and word recognition, sentence and paragraph reading comprehension, word to picture matching and writing on command.

For further explanation and review of the development process, see the materials section and Supplementary material 2 for the lists of subtests of the original version of the SATG, with a detailed descriptions of cultural and linguistic modification.

In short, the content of the SATG assesses language modalities commonly assessed by other aphasia batteries. Tasks and items within the SATG subtest are similar to those in the WAB and BDAE. Therefore, the SATG – in our view – subjectively meets the criteria for content validity.

Discussion

This study presents the preliminary development and pilot psychometric evaluation of the first aphasia test for Gulf Arabic speaking populations. The SATG was developed based on “the aphasia diagnostic informal assessment” by Alzahrani (Alzahrani, 2003), which was developed from the frameworks of two popular Western comprehensive aphasia batteries (BDAE and WAB). The SATG followed a similar theoretical framework as these aphasia batteries, focusing on measures of impairment in different modalities (speaking, understanding, reading and writing). Items and tasks within the SATG were modified to be culturally and linguistically appropriate for Gulf Arabic speaking populations. The SATG was developed for the Gulf Arabic speaking dialect and was normed on healthy adults from Gulf Arabic countries. Only Saudi and Kuwait PWA were recruited in this study due to the access and availability of SLT clinics in other Gulf Arab countries. The SATG was found to be reliable over time and from one clinician to another. The preliminary results also indicate that the SATG has both face and content validity.

The development of the SATG went through three phases and then preliminary norming data was collected from 37 healthy adults with no history of neurological disorder. In this study we failed to match for age and education; principally because of time and resource limitations. The control group were younger and had spent a longer time in education as compared to the PWA group. A future study that provides more substantial normative data, across a breadth of ages and educational levels is needed. For example, we might see more variation and less ceiling effects in a control group that is older and has spent less time in education. However, since the performance of the control group of healthy adults was close to ceiling, beyond providing a view of the influences of age and education, adding more healthy adults will contribute little to the SATG (Howard et al., 2009). For the SATG to be a diagnostically sensitive aphasia test, the SATG psychometric properties should be evaluated using a representative sample of the population under study (Rohde et al., 2018). This means including another control group of people who have suffered a stroke but who

have no language impairment. It should be noted in the case of the SATG that the main reason to include healthy adults at this stage was to obtain the performance of healthy adults in various tasks within the SATG, considering that this is the first aphasia test in the region and no study is available that reports what can be expected from healthy adults in tasks similar to those in the SATG (e.g. picture description). Thus, in this study, the control group provided a means of obtaining scoring criteria for some tasks (e.g. MCA scoring for the picture description, and preliminary normative data for word fluency).

For the SATG, spontaneous speech was examined using a semi-spontaneous task (picture description). The MCA scoring system only checks the presence of a main concept – that is, the information content of speech production. The observational conversational checklist was added to provide more information about the profile of speech production for the PWA. At this stage classifying aphasia syndromes using the SATG scores could not be demonstrated, however, this might be accomplished in future by using a numeric representation of the observational conversational checklist to be computed in addition to the SATG scores across different sections. It is worth noting that spontaneous speech alone is unreliable as a means of classification, as a poor correlation was reported between the site of lesion and the aphasia classification using spontaneous speech (Willmes & Poeck, 1993). Future research that establishes a correlation between the brain lesion site and the performance on the SATG would show whether SATG scores could be distributed according to classification of aphasia syndrome and lesion site.

The preliminary development of the SATG revealed some weaknesses in the test. There was some variation in the total scores of the SATG subtests, then reflected in the overall score. For example, the variation between the scoring of the naming section (scores are out of 56) and the automated speech section (score is out of 2) might suggest that naming tasks are more important than the automated speech to assess PWA language proficiency. Future development may look to take an average or other summary statistic across different sections so that they are equally weighted. The word lists used for naming and auditory comprehension tasks were all selected from high frequency word categories, thus the SATG does not assess whether low frequency words are affected for PWA. Supplementing subtests with low frequency words is necessary, particularly for mild presentations of aphasia. In the automated speech section, the use of reciting a part of the Holy Qur'an caused some PWA participants to report low mood and feelings of depression, as they linked the inability to

recite from the Holy Qur'an with being a 'bad' Muslim. Therefore, future modifications may change the content so as not to include any religious subjects.

In conclusion, this study provides preliminary data for the development of the SATG, its validity and reliability. Considering, this the first aphasia test in the region, it should be noted that the development and the psychometric evaluation of the SATG requires further research. Future plans are to continue the development of the SATG, amending issues highlighted above and completing further tests of psychometric properties. Clinicians are encouraged to report any comments on their use of the SATG or suggestions for improvements directly to the first author.

References

- Abou El-Ella, M. Y., Alloush, T. K., El-Shobary, A. M., El-Dien Hafez, N. G., Abd El-Halim, A. I., & El-Rouby, I. M. (2013). Modification and standardisation of Arabic version of the Comprehensive Aphasia Test. *Aphasiology*, 27(5), 599-614. doi:10.1080/02687038.2013.793282
- Al-Thalaya, Z., Nilipour, R., Sadat Ghoreyshi, Z., Pourshahbaz, A., Nassar, Z., & Younes, M. (2017). Reliability and validity of bedside version of Arabic Diagnostic Aphasia Battery (A-DAB-1) for Lebanese individuals. *Aphasiology*, 32(3), 323-339. doi:10.1080/02687038.2017.1338661
- Alyahya, R. S. W., & Druks, J. (2015). The adaptation of the Object and Action Naming Battery into Saudi Arabic. *Aphasiology*, 30(4), 463-482. doi:10.1080/02687038.2015.1070947
- Alzahrani, S. (2003). *The aphasia diagnostic informal assessment*. Unpublished
- Asseri, A., Alsawiti, F., Alyahya, R., Alanbar, S., & Alharbi, M. (n.d.). *The Arabic Diagnostic Language test (ADLT)*.
- Benson, D. F., & Ardila, A. (1996). *Aphasia: A clinical perspective*: Oxford University Press on Demand.
- Beveridge, M. E. L., & Bak, T. H. (2011). The languages of aphasia research: Bias and diversity. *Aphasiology*, 25(12), 1451-1468. doi:10.1080/02687038.2011.624165
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev*(6), Cd000425. doi:10.1002/14651858.CD000425.pub4
- Byng, S., Kay, J., Edmundson, A., & Scott, C. (1990). Aphasia tests reconsidered. *Aphasiology*, 4(1), 67-91. doi:10.1080/02687039008249055
- Enderby, P. M., Wood, V., & Wade, D. T. (2006). *Frenchay Aphasia Screening Test (FAST)*: John Wiley & Sons.
- Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., . . . Lyrer, P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*, 37(6), 1379-1384. doi:10.1161/01.STR.0000221815.64093.8c
- Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*: John Wiley & Sons.
- Goodglass, H., & Kaplan, E. (1972). *The Assessment of Aphasia and Related Disorders*: Lea & Febiger.
- Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders*: Lippincott Williams & Wilkins.
- Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, 4(2), 627.
- Hassanein, A., El-Tamawy, M., Sallam, T., Hosny, H., Abdel Naseer, M., & El-Fayoumy, N. (2002). Kasr El Aini Arabic Aphasia test (KAAT) simple, standardized, valid, reliable test for Egyptian patients, literate and illiterate. *Egyptian Journal Neurologic Psychiat Neurosurgery*, 39, 381-395.
- Holland, A. L. (1980). *CADL communicative abilities in daily living : a test of functional communication for aphasic adults*. Baltimore: University Park Press.
- Howard, D., Swinburn, K., & Porter, G. (2009). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. *Aphasiology*, 24(1), 56-74. doi:10.1080/02687030802453202

- 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.
- Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test*: Lippincott Williams & Wilkins.
- Kertesz, A. (1982). *Western aphasia battery test manual*: Psychological Corp.
- Khamis-Dakwar, R., & Froud, K. (2012). Aphasia, language, and culture: Arabs in the US. *Aspects of Multilingual Aphasia*, 8, 275-288.
- Khoja, M. A. (2017). A survey of formal and informal assessment procedures used by speech-language pathologists in Saudi Arabia. *Speech, Language and Hearing*, 1-9. doi:10.1080/2050571x.2017.1407620
- Khwaileh, T. A., Mustafawi, E., Howard, D., & Herbert, R. (2016). *An aphasia battery for Qatari/Gulf Arabic*. Paper presented at the Frontiers in Psychology 54th Annual Academy of Aphasia Meeting Conference Abstracts.
- Kong, A. P.-H. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders*, 42(6), 442-464.
- Paradis, M., & El Halees, Y. (1989). *Bilingual Aphasia Test (Jordanian Arabic version)*. Hillsdale, NJ: Lawrence Erlbaum.
- Richardson, J. D., & Dalton, S. G. (2015). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 30(1), 45-73. doi:10.1080/02687038.2015.1057891
- 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. doi:10.1371/journal.pone.0194143
- Rutjes, A. W. S., Reitsma, J. B., Coomarasamy, A., Khan, K. S., & Bossuyt, P. M. M. (2007). Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment - Southampton*, 11(50).
- Saad, M. K., & Ashour, W. (2010). *Arabic text classification using decision trees*. Paper presented at the Proceedings of the 12th international workshop on computer science and information technologies CSIT.
- Shewan, C. M., & Kertesz, A. (1980). Reliability and validity characteristics of the Western Aphasia Battery (WAB). *Journal of Speech Hearing Disorders*, 45(3), 308-324.
- Simera, I., Altman, D. G., Moher, D., Schulz, K. F., & Hoey, J. (2008). Guidelines for reporting health research: the EQUATOR network's survey of guideline authors. *PLoS Medicine*, 5(6).
- Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*: Psychology Press.
- Wagenaar, E., Snow, C., & Prins, R. (1975). Spontaneous speech of aphasic patients: A psycholinguistic analysis. *Brain and Language*, 2, 281-303.
- Willmes, K., & Poeck, K. (1993). To what extent can aphasic syndromes be localized? *Brain*, 116(6), 1527-1540.
- Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS One*, 13(2), e0192773. doi:10.1371/journal.pone.0192773

Appendix 1: The manual of the Short Aphasia test for Gulf Arabic speaker

The Short Aphasia Test for Gulf Arabic speaker (SATG)

The final version (second draft) of the SATG consists of six sections that assess different language skills: semi-spontaneous speech, auditory comprehension, repetition, naming, automatic speech, recitation, reading and writing. In total, the SATG includes 14 subtests distributed in six sections. Below we will describe each subtest briefly, refer to Supplementary material 6 for the cover and scoring sheet of the final version of the SATG and Supplementary material 7 for the SATG material book.

Section I: Semi-spontaneous speech (maximum score = 12 points)

Purpose: to obtain a semi-spontaneous speech sample using complex picture description. It is intended to measure expressive speech that may be more relevant for functional communication and to provide a baseline and outcome measure during therapy (Richardson & Dalton, 2015).

Subtest:

- A. Picture description (living room gathering–Figure 1): This picture was developed by a group of Saudi SLTs as part of clinical research to develop the comprehensive aphasia test ADLT (Asseri et al., n.d.). **Scoring criteria:** The scoring criteria for the picture description task were designed to evaluate the patient’s production using the main concept analysis (MCA) approach based on a proposition-level measure. The MCA measures the ability of the patient to provide the overall gist/essential information of the stimulus picture (Kong, 2009; Richardson & Dalton, 2015). MCA is informative, easy, objective and reliable scoring approach. Refer to Supplementary material 8 for details on the main concept analysis (MCA) approach for scoring picture description and adapting MCA to the SATG picture description task.

Section II: Auditory comprehension (maximum score = 35 points)

Purpose: to measure verbal auditory comprehension of words from multiple semantic categories, sentences, and questions of increasing complexity (Shewan & Kertesz, 1980).

Subtests:

- A.** Word auditory comprehension (maximum score = 16 points) in this subtest, the patient is asked to select a target picture from a set of 4 pictures corresponding to a spoken word. Sixteen words from three categories (nouns, verbs and colours) were selected according to their high frequency and everyday familiarity based on focus group report and a list produced by (Alyahya & Druks, 2015). Pictures were drawn especially for the SATG and judged by the focus group and pilot data for their clarity. Please refer to Supplementary material 9 for the list of pictures in the SATG. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response.
- B.** Simple and complex commands: (maximum score = 8 points) in this subtest, the patient is asked to perform eight simple, semi-complex, and complex commands. These range from one step to multiple steps within a specific order (3 simple commands that required one simple action, 3 semi-complex commands that required two actions, and 2 complex commands that required 3 actions). All commands were designed to avoid using objects, in order to be suitable for the digital format of the SATG. **Scoring criteria:** patients score 1 point for each fully correct response and 0 points for incomplete execution, incorrect or absent responses.
- C.** Yes/No questions: (maximum score = 11 points) in this subtest, the patient is asked to answer eleven questions using either yes or no. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response. Only a verbal response is accepted. If the patient uses gesture or opens/closes their eyes to respond, the examiner can mention the response type and score this separately

Section III: Repetition (maximum score = 12 points)

Purpose: to test repetition of single words and sentences of increasing length and complexity. Repetition measures expressive and phonological processing output and short-term auditory memory.

Subtest:

- A. Word and sentence repetition: the patient is asked to repeat 12 items after the examiner. It starts with one word (one-syllable) and proceeds to multiple-syllabic words and sentences of increasing length and complexity (1-syllable words to 6-7 word sentences of 17-18 syllables). **Scoring criteria:** patients score 1 point for correct production of a word/sentence and 0 points for each incorrect or absent response. The examiner should determine the type of incorrect response and code any errors (circumlocution, or phonemic, semantic, or neologistic paraphasia). See Supplementary material 6.

Section IV: Naming (maximum score = 56 points)

Purpose: lexical access and retrieval during spoken word production.

Subtest:

- A. Picture word naming: (maximum score = 21 points) in this subtest, the patient is asked to name 21 words from four categories (nouns, verbs, body parts and colours). Similar to the word auditory comprehension subtest, words were selected according to their high frequency and everyday familiarity based on focus group reports and a list produced by (Alyahya & Druks, 2015). See Supplementary material 9 for the list of pictures. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response. Examiners refer to the error codes table to describe the type of incorrect responses.
- B. Sentence completion: (maximum score = 11 points) in this subtest, the patient is asked to complete eleven sentences presented by the examiner using one word only. Sentences increase in length and complexity. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response. Examiners refer to the error codes table to describe the type of incorrect responses.
- C. Word fluency: (maximum score = 16 points) in this subtest, the patient is asked to name as many animals as she/he can within one minute. After focus group reports and results from the control healthy adult and pilot data, the animal category was chosen over the clothes category for familiarity. **Scoring criteria:** patients score 1 point for each correct animal named, and 0 points for each incorrect or absent response, with the maximum score (16). Examiners refer to the error codes table to describe the type of incorrect responses.

- D. Responsive naming:** (maximum score = 8 points) in this subtest, the patient is asked to answer eight questions asked by the examiner using one word only. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response. Examiners refer to the error codes table to describe the type of incorrect responses.

Section V: Automatic speech and recitations (maximum score = 2 points)

Purpose: to measure the ability to produce an automatized sequence of numbers and recitation of passages which are culturally appropriate and should be familiar to the patient.

Subtest:

- A. Automatic speech** (maximum score = 1 point): in this subtest, the patient is asked to count from 1-20. **Scoring criteria:** patients score 1 point for a correct complete response and 0 points for an incomplete and/or incorrect or absent response.
- B. Recitations** (maximum score = 1 point): in this subtest, the patient is asked to recite part of the holy Qur'an. **Scoring criteria:** patients score 1 point for a correct complete response and 0 points for an incomplete and/or incorrect or absent response.

Section VI: Reading and Writing (maximum score = 31 points)

Purpose: to measure the matching of spoken to written forms, and reading comprehension of words, sentences, and a paragraph. Also, writing ability (i.e. the writing of the patient's own name) is measured. These tests target the decoding of written items, and orthographic, semantic and phonological processing.

Subtest:

- A. Spoken letter/word to written letter/word choice matching** (maximum score = 14 points) in this subtest, the patient is asked to select written letter/word from set of 4-written letter/words corresponding to verbal stimulus. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response.

- B.** Reading comprehension (maximum score = 7 points). The patient is asked to read either aloud or silently sentences and a paragraph, which increase in length and complexity. They then choose from four written (word) options the correct answer that completes the sentence. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response.
- C.** Picture to written word choice matching (maximum score = 8 points). The patient is asked to look at a picture presented and select the correct written word from a choice of four. **Scoring criteria:** patients score 1 point for each correct response and 0 points for each incorrect or absent response.
- D.** Writing on demand: (maximum score = 2 points) in this subtest, the patient is asked to write his/her full name (first and family name). If the patient fails, then the examiner asks the patient to copy his/her name. **Scoring criteria:** patients score 2 points for writing their full name correctly, 1 point for missing one of the names or misspelling either one of the names and 0 points for copying, or for an incorrect response (i.e. writing different name or misspelling both names) or no response. The examiner should document the reason for giving 0 points.

Conversational observation checklist:

This checklist was added to provide very brief insight into some aspects of the spontaneous conversational interaction of PWA while the SATG is conducted. Also, it was added to give a summary of the speech characteristics and content of the PWA's production. The main characteristics of this checklist were taken from both the WAB (fluency of speech) and BDAE (rating scale profile of speech characteristics). It was designed to be straightforward and easy to complete after the SATG has been completed. The observational checklist contains six characteristics (speech, articulation, meaning, grammar, error recognition, response to cuing and paraphasia type). Please refer to Supplementary material 6 for the conversational observation checklist within the cover of the final version of the SATG.