The Screening Test for Aphasia and Dysarthria (STAD) for Patients with Neurological Communicative Disorders: A Large-Scale, Multicenter Validation Study in Japan

Kentaro Araki^{a, b} Yoshiyuki Hirano^{a, c} Machiko Kozono^d Junko Fujitani^e
Eiji Shimizu^{a, c}

^aResearch Center for Child Mental Development, Chiba University, Chiba, Japan; ^bDepartment of Rehabilitation, Mitsuwadai General Hospital, Chiba, Japan; ^cUnited Graduate School of Child Development, Osaka University, Kanazawa University, Hamamatsu University School of Medicine, Chiba University and University of Fukui, Chiba, Japan; ^dDepartment of Rehabilitation, Kumamoto Health Science University, Kumamoto, Japan; ^eDepartment of Rehabilitation, National Center for Global Health and Medicine, Tokyo, Japan

Keywords

Aphasia · Brain injury · Cognitive dysfunction · Communicative disorders · Dysarthria

Abstract

Introduction: Evaluation of multiple domains, such as language, articulation, and cognitive function, is frequently required in neurological communicative disorders. The purpose of this study was to investigate the performance of a 10-min screening scale for estimating aphasia, dysarthria, and cognitive dysfunction using a multicenter, large-sized consecutive series. **Methods:** We conducted a multicenter validation study that included 314 patients with brain injury between February 1 and June 31, 2018, from 20 medical centers across Japan. The Screening Test for Aphasia and Dysarthria (STAD) was developed in Japan in 2009, and a previous smaller-scale retrospective study established its high to moderate validity. All patients had undergone the STAD, and 212 of them underwent the Western Aphasia Battery or Assessment of Motor Speech for Dysarthria. The effect size on all 29 items and receiver operating curves of 3 sections of the

STAD were analyzed based on external criteria, which were decided considering the clinical diagnosis of aphasia, dysarthria, and cognitive dysfunction. Correlations between the STAD and reference tests were calculated. **Results:** The *phi* coefficients of 23 out of 29 items exceeded the moderate effect size of 0.3 toward the targeted disorder. Overall, there was a good balance between sensitivity (82–92%) and specificity (77–78%), with moderate to large positive and negative likelihood ratios (3.7–4.19 and 0.1–0.23). The Pearson's *r* between the verbal section and Western Aphasia Battery Aphasia Quotient, the articulation section and Assessment of Motor Speech for Dysarthria, and the nonverbal section and Western Aphasia Battery Nonlinguistic Skills were 0.89, 0.70, and 0.79, respectively. **Conclusion:** We demonstrated that the STAD has acceptable content and concurrent validity for the assessment of communicative function in patients with brain injury. This short screening tool can be useful in specific contexts, such as in early bedside investigations, to obtain a quick summary of communicative function prior to the administration of other tests, and in cases where more in-depth testing is not feasible.

© 2021 The Author(s).
Published by S. Karger AG, Basel

Introduction

Communication disorders are an established consequence of brain injury. Aphasia is an acquired selective impairment of language modalities and functions resulting from brain lesions in the language-dominant hemisphere that affects the person's communicative and social functioning, the quality of life of the patients, and their relatives and caregivers [1]. Approximately one-third of those who experience a stroke develop aphasia [2, 3]. Aphasia is typically a consequence of damage to the left hemisphere, in which the area of the brain responsible for language function is usually situated in right-handed people [4]. Dysarthria is a collective name for a group of neurological speech disorders that reflect abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for the breathing, phonatory, resonatory, articulatory, or prosodic aspects of speech production [5]. The epidemiology of dysarthria has changed recently, with an increased number of diagnoses (likely due to aging populations) [5]. Dysarthria affects approximately 20–30% of stroke survivors [6, 7] and 10–60% of those who survive traumatic brain injury (TBI) [8]. A prospective study indicated that the cause of dysarthria due to a single episode of ischemic stroke was located in the motor cortex (14.5%), striatocapsular lesion (46.8%), base of the pons (24.2%), and cerebellum (14.5%) [9].

In addition to aphasia and dysarthria, cognitive-communicative disorders often affect patients after brain injury; the estimated worldwide prevalence of cognitive-communicative impairments after the first episode of stroke ranges from 39% [10] to 77% [11, 12]. Cognitive skills include processes such as attention, memory, reasoning, problem-solving, and executive functioning. Typically, dysfunctions of the prefrontal lobe (executive function), right hemisphere (unilateral neglect and general attention), Papez circuit (episodic memory), and diffuse lesions, as in the case of TBI, lead to the development of cognitive-communicative disorders. However, as these functions are not uniform, and their presentation and severity depend on the disrupted brain networks, the symptoms attributed to those dysfunctions are not restricted to damage solely to the aforementioned areas [13]; injury to any part of the brain may cause cognitive dysfunction.

It has repeatedly been suggested that the treatment of communicative disorders should be initiated as soon as possible after brain injury [14–16]. Early involvement in appropriate, intensive therapy coincides with patterns of neural recovery [17] and serves to increase the opportunities for the development of coping strategies for commu-

nication deficits, thereby reducing patient isolation and promoting participation in the rehabilitation process [18, 19]. Therefore, early identification and diagnosis of communicative deficits is an important step toward maximizing the positive effects of rehabilitation.

Despite the usefulness of standardized cognitive-linguistic assessments, the time required to administer them often renders them unsuitable [20, 21]. This is especially true for patients in a state of confusion, which is often observed in cases of acute stroke or unstable status (also observed in acute phases) [20, 22]. Although communicative screening instruments should never be used alone to establish a diagnosis, they do contribute toward identifying the need for additional assessments [23]. Communicative screening tests may help track a patient's disorder longitudinally by analyzing the patients' early clinical course [24–26], identify early management decisions [27, 28], test patients who may not tolerate long evaluation processes [29, 30], provide better personalized advice to families [29], and contribute to diagnostic accuracy [31].

Three guidelines (Veterans Administration/Department of Defense, Scottish Intercollegiate Guideline Network, and Canadian Best Practice Recommendations) recommend that all aspects of communication and cognition, which include speech, language, attention, memory, executive functions, and visual-spatial abilities, should be assessed to determine rehabilitation needs [26, 27, 32, 33]. However, except for the scale of cognitive and communicative ability for neurorehabilitation (SCCAN) [34], to our knowledge, there was no mention concerning multiple-domain screening in the recent literature. A limitation of existing screening tools is that they are all domain-specific in estimating an individual's overall communicative ability. For example, when a patient with brain injury receives a solely language-focused screening test, the patient's performance on the test is negatively affected by the presence of dysarthria [23, 35] and cognitive disorders, especially visual neglect and attention deficits [29, 36–38], despite having normal language function. As dysarthria and cognitive disorders are frequently observed in brain injury, this raises questions regarding the existing aphasia screening tests from the viewpoint of false positive responses caused by the presence of other disorders, which may be overlooked.

The Screening Test for Aphasia and Dysarthria (STAD) was designed in Japan in 2009 to provide an overview of the cognitive and communicative abilities of patients with brain injury, enable evaluation with a moderate to high degree of reliability and validity, and permit rapid test administration (approximately 10 min) [39]. The

STAD was designed to be suitable for bedside, home, or examination room administration by speech-language pathologists (SLPs), psychiatrists, or neurologists. One way to ascertain the overall picture of the communicative function across multiple areas is to simultaneously test all the domains of communicative function [40, 41]. Therefore, including 3 test sections that respond specifically to the 3 different disorders (aphasia, dysarthria, and cognitive dysfunction) facilitates the estimation of the pivotal domain that inhibits patient communication. Thus, the STAD focuses on language, articulation, and cognitive abilities measured separately in verbal, articulation, and nonverbal sections, respectively.

In a previous, smaller-scale retrospective study investigating the STAD's psychometric properties, Araki et al. [42] administered the screening test to 45 patients with stroke who were suspected of having communication disorders within 1 month after stroke onset. The average time required for administration was 9 min 48 s. Another study published in 2019, conducted on healthy individuals, provided normative data of 222 native Japanese individuals aged >50 years [43]. The mean scores (standard deviation [SD]) on verbal, articulatory, and nonverbal tests were as follows: 15.8 (0.64) out of 16, 6.9 (0.44) out of 7, and 5.9 (0.36) out of 6, respectively. When a value of ± 1.5 SD is taken as the cutoff, in order to encompass 86.6% (43.3% on each side) of the data, as indicated by the empirical rule, the cutoffs for the verbal, articulatory, and nonverbal scores are 14.8, 6.2, and 5.3, respectively; scores lower than these cutoffs would suggest abnormality. A concurrent validity study conducted in 2018 on 48 stroke patients showed a close correlation between the patients' scores on the 3 STAD sections and referenced standard tests [39]. The correlation coefficient between the STAD verbal section and the Western Aphasia Battery (WAB) aphasia quotient (AQ) was 0.90; that of the STAD articulation section with the Assessment of Motor Speech for Dysarthria (AMSD), which is widely used in Japan and will be discussed further in the Materials and Methods section, was 0.70; and that of the STAD nonverbal section with WAB nonlinguistic skills (WAB NLS) was 0.79. After considering all these factors, the Japanese version of the STAD met the criteria for test standardization in Japan in 2018.

In this study, we aimed to assess the validity of the STAD on a larger scale, with a consecutive series that included over 300 patients with stroke and with physicians' prescriptions, selected from 20 multicenter studies in Japan between February 1 and June 31, 2018. We verified the validity of the STAD in 3 ways. First, we assessed item-

Fig. 1. Flowchart of the recruitment procedure.

level validity and calculated the effectiveness of all the STAD items in detecting the presence or absence of aphasia, dysarthria, and cognitive dysfunction. We expected to observe stronger effect sizes between the items belonging to the verbal section and aphasia, those in the articulation section and dysarthria, and those in the nonverbal section and cognitive dysfunction. Especially, when an item correctly correlates with the target disorder, the effect size of the items related to the target disorder may increase. Second, we calculated the correlations between the patients' scores on the STAD and those on different reference tests to examine concurrent validity. We expected significant correlations between the STAD and the other targeted measures. Third, we examined the test's sensitivity and specificity to determine whether the STAD could differentiate between patients with brain injury with and without communicative disorders. We hypothesized that the sensitivity and specificity of the STAD would be sufficient for screening purposes.

Materials and Methods

Participants

The study participants were consecutively selected from 20 medical centers in Japan (7 patients from Kanto, 7 from Kyusyu, 2 from Kansai, 2 from Tyugoku, and 2 from Tyugoku). We included patients aged >18 years who were native Japanese speakers with

brain injuries (including cerebral infarction, cerebral hemorrhage, subarachnoid hemorrhage, TBI, and subdural or epidural hematoma) with an onset within the previous 180 days. Speech and language rehabilitation was prescribed by physicians between February 1 and June 31, 2018. We excluded participants with any of the following characteristics: (1) medically unstable (e.g., high blood pressure or heart rate, low oxygen saturation, and/or progression or exacerbation of symptoms) to avoid the occurrence of adverse events during the study; (2) degenerative diseases; (3) disturbance in consciousness (e.g., those who could not open their eyes by themselves); (4) severe visual or auditory problems hindering their ability to undergo the examination; or (5) consent for participation not provided. This study adhered to the Declaration of Helsinki guidelines for experiments involving humans. The Ethics Review Committee of Chiba University approved this study (ID: 2627). Written informed consent was obtained from all participants. In cases where a patient could not provide informed consent because of a severe communicative disorder, written consent was obtained from their close relatives. During the study period, 591 consecutive, potentially eligible participants were recruited (Fig. 1). Of these, 186 were excluded because of their condition(s): 75 (40%), degenerative diseases; 47 (25%), disturbance in consciousness; 32 (17%), medical instability; 17 (9%), transient ischemic attack without neurological signs on admission; and 15 (8%), severe visual or auditory problems. Seventy-three patients declined to participate, and 18 were excluded for other reasons (inability to administer the STAD due to early discharge, SLP's judgment of inappropriateness due to diseases not indicated in the exclusion criteria, such as locked-in syndrome and severe rheumatism, or end of the study period). Finally, 277 patients were excluded from the 591 potentially eligible participants, and 314 patients were enrolled.

The demographic and clinical profiles of the study participants are summarized in Table 1 and seem to reflect the typical characteristics of Japanese patients with brain injury when compared with another larger-scale study [44]. This is important because our study criteria excluded cases of severe illness (i.e., patients with modified Rankin Scale [mRS] = 6) and those with mRS = 0 (no symptoms) or 1 (no significant disability) who would not have been prescribed speech and language rehabilitation.

In total, 212 patients completed the STAD. Within the following 2 weeks, the reference standard tests of the Japanese version of the WAB [45, 46] or the AMSD, a test widely used in Japan to assess dysarthria, were administered [47]. Their descriptive data were as follows: males, 126 (59%); mean age, 72.4 ± 13.4 ; infarction, 134 (63.2%); hemorrhage, 47 (22.2%); subarachnoid hemorrhage, 7 (3.3%); trauma, 7; (3.3%); average mRS, 3.2 ± 1.2 . According to our study protocol, described later, 102 patients underwent the STAD but not the reference test. Among them, 60 patients did not meet the inclusion criteria: 32 (53%) had neither aphasia, dysarthria, or cognitive dysfunction (their type of brain injury was a transient ischemic attack or a very small lacunar infarction); 24 (40%) felt that the burden of the reference test was too high; and 4 (7%) had stroke deterioration. Moreover, 18 patients declined to participate, and 24 were excluded for other reasons.

Procedure

During the study period, the STAD was administered individually to eligible participants prior to the reference standard tests. It was conducted at the bedside, in the rehabilitation room, or in the dayroom at the patients' hospitals. To confirm the pres-

Table 1. Demographic and clinical profiles of the study participants

Characteristic	n (valid %)
Sex	
Woman	133 (42.4)
Man	181 (57.6)
Age, years	
<65	65 (20.1)
65–74	82 (26.1)
75–84	110 (35.0)
≥85	57 (18.2)
Mean ± SD, range	72.7±12.9
Disease duration (days)	
<10	99 (31.5)
10–29	61 (19.4)
30–59	55 (17.8)
60–180	96 (30.6)
Undetectable ^a	2 (0.6)
Mean ± SD, range	45.7±47.8
Etiology	
Infarction	204 (65.0)
Hemorrhage	67 (21.3)
Subarachnoid hemorrhage	12 (3.8)
Trauma	11 (3.5)
Others	20 (5.7)
Brain lesion	
Left hemisphere	132 (42.0)
Right hemisphere	101 (32.2)
Bilateral hemisphere	13 (4.1)
Cerebellum/brainstem	36 (11.5)
Others ^b	26 (8.3)
Missing ^c	6 (1.9)
Brain injury severity	
mRS 0	7 (2.2)
mRS 1	34 (10.8)
mRS 2	71 (22.6)
mRS 3	54 (17.2)
mRS 4	110 (35.0)
mRS 5	38 (12.1)
mRS 6	0 (0)
Mean ± SD	3.1±1.3

mRS, Modified Rankin Scale; SD, standard deviation. ^a Undetectable onset day of chronic subdural hematoma patients who had not undergone surgery. ^b Atypical lesion (e.g., hemispheric and cerebellum/brainstem simultaneous lesions caused by multiple cerebral infarctions or lesions without focal sites caused by hydrocephalus, subarachnoid hemorrhage, subdural hematoma, or epilepsy. ^c Containing data inadequacy.

ence or absence of aphasia, dysarthria, and cognitive dysfunction, licensed SLPs, approved by the Japanese Ministry of Health to practice in Japan, used standardized neurological communication examinations commonly available in Japan along with findings from brain imaging and clinical observations of the patients. The neurological communication examinations included the Standard Language Test of Aphasia [48], Token Test [49], and WAB

[46] for aphasia; AMSD for dysarthria [47]; Japanese Mini Mental State Examination [50], Kohs Block Design Test [51], Raven’s Colored Progressive Matrices [52], Rivermead Behavioral Memory Test [53], Trail making Test [54], Wechsler Adult Intelligence Scale-III [55], and/or Wisconsin Card Sorting Task for cognitive dysfunction [56].

Aphasia and dysarthria were diagnosed based on the definitions of the disorders described at the beginning of this article. Cognitive dysfunction was considered to be present when the patient presented with one or more of the following symptoms: attention deficit, visuospatial neglect, ideomotor apraxia, constructional apraxia, executive dysfunction, or memory dysfunction [42]. A total of 58 SLPs with an average of 7.6 years of experience (SD ± 5.7) were involved in this study (sex, 39 women, 18 men). Fifty-seven of them were full-time practitioners, and one was a part-time practitioner.

In total, 92 (29%), 154 (49%), and 179 (57%) patients exhibited aphasia, dysarthria, and cognitive dysfunction, respectively. The patients were classified into 8 specific disorder classes based on the presence or absence of each of these symptoms. Table 2 shows this concept and the number of patients in each category. Patient characteristics were used as external criteria in the item response and diagnostic accuracy analyses.

Generally, we attempted to complete the reference standard tests within 2 weeks of STAD administration. They were individually conducted by licensed SLPs in an exam room designated for speech, language, and/or hearing therapy in the hospital. We investigated the extent of correlation with 3 reference standard tests, as each test measures a different aspect of communicative ability. As reference standard tests, we chose the WAB AQ, the AMSD, and the WAB NLS, as these tests have already been standardized, are commonly used in Japan, and quantitatively evaluate the patients’ communicative abilities and disorder severity for aphasia, dysarthria, and cognitive dysfunction.

In this study, blinding procedures, wherein, for example, a tester unaware of the results of the index test would carry out the reference test, were not implemented, partly because several SLPs belonged to hospitals that had only one SLP. Ideally, all 3 tests could have been performed on each patient; however, for ethical reasons, we performed only one test to avoid placing too much burden on the patients. Therefore, we created a protocol to pair the main disorder manifested by the patient to one of the reference tests: the WAB AQ, AMSD, and WAB NLS for patients with aphasia, dysarthria, and cognitive dysfunction, respectively.

Measures

Index Measure

STAD. The STAD comprises 3 sections and 29 items in total. Each item was scored 1 point. A preliminary English version of the STAD and a plain glossary to standardize the examination procedures and scoring criteria can be shared upon request. Anonymized data for the primary analyses presented in this report are also available upon reasonable request.

The STAD verbal section contains 16 items in total, which estimate aphasia: patient name; obey command 1, 2, and 3; word repetition 1, 2, and 3; automatic speech; object naming 1, 2, and 3; name writing 1 and 2; dictation 1, 2, and 3. The instructions for each examination and each item’s difficulty are shown in online supplementary Material S1 (for all online suppl. material, see www.karger.com/doi/10.1159/000519381). A ranking of items from

Table 2. Patient categorization based on disorder-presence/-absence across the 3 types of disorders and the corresponding patient frequency (number of patients [%])

Type	Aphasia	Dysarthria	Cognitive dysfunction	Patients (n = 314)
A	✓	✓	✓	26 (8.3)
B	✓	✓	+	7 (2.2)
C	✓	+	✓	37 (12.8)
D	✓	+	+	22 (7.0)
E	+	✓	✓	67 (21.3)
F	+	✓	+	54 (17.2)
G	+	+	✓	49 (15.6)
H	+	+	+	52 (16.6)

There were 8 categories according to the presence or absence of aphasia, dysarthria, and cognitive-related communicative dysfunction (cognitive dysfunction). Aphasia was present in patient types A–D (n = 92) and absent in patient types E–H (n = 222). Dysarthria was present in patient types A, B, E, and F (n = 154) and absent in patient types C, D, G, and H (n = 160). Cognitive dysfunction was present in patient types A, C, E, and G (n = 179) and absent in patient types B, D, F, and H (n = 135).

easy to difficult was created based on the distribution of the rate of correct answers in 314 cases. For instance, items, such as patient name, where 285 (90.8%) out of 314 patients with stroke answered correctly, were classified at the top (easy); in contrast, items, such as dictation 3, where 163 (51.9%) patients answered correctly, were placed at the bottom of the ranking and classified as difficult. The STAD-articulation section estimated dysarthria based on 7 items: oral movement 1, 2, 3, and 4, and diadochokinesis 1, 2, and 3. The order of difficulty of the items in the articulation section presented in online supplementary Table A1 ranges from easier ones (with a correct answer rate of 80%) to more difficult ones (such as oral diadochokinesis 3, with a correct answer rate of 55.4%). The nonverbal section includes 6 items that estimate cognitive dysfunction: eye contact, orientation, imitation 1 and 2, and visual construction 1 and 2. The eye contact item aims to evaluate attention deficit and unilateral spatial neglect. The orientation item evaluates memory deficits. The imitation item estimates attention deficit, unilateral spatial neglect, ideomotor apraxia, and limb kinetic apraxia. The visual construction item estimates attention deficit, constructional apraxia, and unilateral spatial neglect. The level of difficulty in the nonverbal section of the STAD also ranged from easy items, such as eye contact (95.5%), to more difficult items, such as cube drawing (53.2%).

The difficulty of the different items is related to the severity (severe to mild) of the communication disorder. Therefore, patients who responded correctly to items with high difficulty levels, such as dictation 3 in the verbal section, oral diadochokinesis 3 in the articulation section, or cube drawing in the nonverbal section, had a mild disorder or none at all. Conversely, patients who responded incorrectly to items with low difficulty levels, such as obey command 1, eye contact, or imitation 1, had the most severe disorder levels. However, several STAD items presented a skewed distribution with very high rates of correct responses. We noted this

Table 3. Example of 2 binary diagnoses in a 2-binary examination response table

	Disorder		Total
	present	absent	
Examination			
Fail	<i>a</i>	<i>b</i>	<i>r</i> ₁
Pass	<i>c</i>	<i>d</i>	<i>r</i> ₂
Total	<i>c</i> ₁	<i>c</i> ₂	<i>N</i>

The table shows binary data of disorder present/absent in the columns and binary data of wrong/correct answers in the rows. The number of cases corresponding to each cell are represented by *a*, *b*, *c*, *d*. The totals of rows one and two are *r*₁ and *r*₂, respectively, and those of columns one and 2 are *c*₁ and *c*₂, respectively. The *phi* coefficient value approximates 1 when *a* and *d* are higher, approximates −1 when *b* and *c* are higher, and approximates 0 when there is no bias in *a*–*d*.

skewed distribution in the following items in a later analysis: obey command 1 (correct answer rate, 94.9%), eye contact (95.5%), and imitation 1 (93.0%).

Reference Standards for Communicative Impairments

Reference Standard I. The WAB is a standard scale widely used for evaluating aphasia [46]. Sugishita [45] conducted a validation study of the Japanese version of the WAB, analyzing its performance in 203 aphasia patients and 32 healthy adult participants. The study confirmed that the responses in the Japanese and English versions of the WAB were very similar. In 1986, the Japanese version of the WAB, which is now widely used in Japan, was released [45]. The WAB assesses linguistic and nonlinguistic skills. In this study, patients with aphasia were administered the Japanese version of the WAB, and we calculated the WAB AQ, which ranges from 0 to 100 (severe to good condition). Patients with a cognitive disorder were administered Part VIII of the WAB, which contains constructional, visuospatial, and calculation items that include drawing, block design, calculation, and Raven's colored progressive matrices tasks. Part VIII of the WAB estimates the patient's nonlinguistic skills (WAB NLS), with a score ranging from 0 to 10 (severe to good condition).

Reference Standard II. Patients with dysarthria underwent the AMSD, which contains 29 items [47] and is widely used in Japan for the standardized evaluation of dysarthria. Each item consists of 4 points, with a cutoff of 3 points; a score ≤3 points indicates a deficit. We summed the subtest results to obtain an indication of the overall severity of dysarthria, with 29 being the best condition and 0 being the worst.

Statistical Analysis

Item-Level Analysis

To assess how each item on the STAD relates to aphasia, dysarthria, and cognitive dysfunction, we evaluated the effectiveness of each item using the *phi* coefficient. The latter is also called the 4-fold point correlation coefficient and can be explained through a 2-by-2 cross table (Table 3). As the *phi* coefficient is defined by

Table 4. Effect size of the “patient name” item according to the presence of aphasia, dysarthria, or cognitive dysfunction

		Aphasia		Total
		present	absent	
<i>phi</i> = 0.45 (medium effect)				
Name				
Fail	27	2	29	
Pass	65	220	285	
Total	92	222	314	

		Dysarthria		Total
		present	absent	
<i>phi</i> = 0.03 (no effect)				
Name				
Fail	13	16	29	
Pass	141	144	285	
Total	154	160	314	

		Cognitive dysfunction		Total
		present	absent	
<i>phi</i> = 0.14 (small effect)				
Name				
Fail	23	6	29	
Pass	156	129	285	
Total	179	135	314	

applying the Pearson correlation coefficient to a 2-by-2 table, the relationship between the column (in this case, disease “present” or “absent”) and row categories (item “fail” or “pass”) could be revealed by the *phi* coefficient as follows:

$$\Phi = \frac{ad - bc}{\sqrt{r_1 r_2 c_1 c_2}} \quad (1)$$

The *phi* coefficient and Pearson correlation coefficient vary from −1 to 1, and the effect sizes of the absolute *phi* coefficient value correspond to 0.1, 0.3, and 0.5, indicating weak, moderate, and strong effects, respectively [57].

Table 4 shows the “correctness” of the patients’ answers to the patient name item according to the presence or absence of the disorders. The patient name item, which belongs to the verbal section of the STAD for the evaluation of aphasia, was registered as correct when the patients could say their first and last names, and as wrong when they could not. All 3 tables, which are part of Table 4, have the same number of patients (29 who failed and 285 who passed), but the bias of correct and incorrect proportions on 4 cells divided by row and column factors was different between the disorder types.

The patients in the aphasia group tended to make mistakes in the item more often (aphasia was present in 27 [93%] cases and absent in 2 cases [7%]). Among the item pass group, aphasia was absent in 220 cases (77%) and present in 65 (23%) cases. Most

Table 5. General clinical and functional characteristics of patients with brain injury and their performance on the index test based on the presence of aphasia, dysarthria, or cognitive dysfunction

Characteristic	Brain injury (n = 314)		effect size	p value
	aphasia			
	present (n = 92)	absent (n = 222)		
Disease duration, days ^a	45.4 (45.6)	46.6 (52.4)	0.02	0.839
mRS	3.3 (1.2)	3.0 (1.3)	0.25	0.030
Lesion site, % ^b			0.47	<0.001
Left hemisphere	72 (76.6)	65 (29.5)		
Right hemisphere	9 (9.6)	87 (39.5)		
Bilateral hemisphere	2 (2.1)	11 (5.0)		
Cerebellum/brainstem	0 (0.0)	36 (16.4)		
Others	8 (8.5)	18 (8.2)		
Index test ^c				
STAD verbal	56.4 (29.4)	93.0 (13.0)	1.90	<0.001
STAD-articulation	67.5 (36.9)	75.5 (28.2)	0.26	0.040
STAD-nonverbal	67.5 (31.1)	81.4 (22.8)	0.55	<0.001

Characteristic	dysarthria		effect size	p value
	present (n = 154)	absent (n =160)		
Disease duration, days ^a	42.5 (44.7)	48.9 (50.3)	0.07	0.235
mRS	3.3 (1.2)	2.9 (1.4)	0.38	0.001
Lesion site, % ^b			0.22	0.005
Left hemisphere	64 (41.6)	73 (47.4)		
Right hemisphere	53 (34.4)	43 (27.9)		
Bilateral hemisphere	5 (3.2)	8 (5.2)		
Cerebellum/brainstem	23 (14.9)	13 (8.4)		
Others	5 (3.2)	21 (13.6)		
Index test ^c				
STAD-verbal	82.4 (26.4)	82.1 (24.6)	0.01	0.913
STAD-articulation	55.3 (28.2)	90.4 (23.4)	1.36	<0.001
STAD-nonverbal	79.4 (26.3)	75.2 (26.1)	0.16	0.154

Characteristic	cognitive dysfunction		effect size	p value
	present (n = 179)	absent (n = 135)		
Disease duration, days ^a	53.7 (48.1)	35.0 (44.9)	0.40	<0.001
mRS	3.6 (1.1)	2.4 (1.2)	1.06	<0.001
Lesion site, % ^b			0.27	<0.001
Left hemisphere	71 (39.7)	66 (48.9)		
Right hemisphere	67 (37.4)	29 (21.5)		
Bilateral hemisphere	11 (6.1)	2 (1.5)		
Cerebellum/brainstem	11 (6.1)	25 (18.5)		
Others	13 (7.3)	13 (9.6)		
Index test ^c				
STAD-verbal	75.4 (27.8)	91.3 (18.6)	0.66	<0.001
STAD-articulation	68.1 (33.6)	80.0 (26.4)	0.39	<0.001
STAD-nonverbal	63.7 (26.8)	95.5 (8.9)	1.51	<0.001

Data are reported as means (SDs) and proportions (%). Percentages indicate the proportions between participant groups. Differences in means were calculated by Student *t* and Cohen's *d* tests for significance and effect size, respectively. Differences in proportion were calculated by χ^2 and Cramer's *V* for significance and effect size, respectively. ^a *n* = 312 (undetectable onset day of chronic subdural hematoma patients who had not undergone surgery). ^b *n* = 307 (containing data inadequacy). ^c Correctness on each STAD section. mRS, modified Rankin Scale; STAD, Screening Test for Aphasia and Dysarthria.

cases fell on the diagonal cells in the table, and the *phi* coefficient value was 0.45, showing a moderate effect size between aphasia and the patient name item. Conversely, the middle table showed no correlation between dysarthria and the patient name item, as most of the data fell off the diagonal (in the item fail group, dysarthria was present and absent in 13 [45%] and 16 [55%] cases, respectively, while in the item pass group, dysarthria was present and absent in 141 [49%] and 144 [51%] cases, respectively; the *phi* coefficient was 0.03). The bottom table for cognitive dysfunction showed roughly equal proportions and a small effect of the *phi* coefficient (0.14). A high coefficient was observed for aphasia and a low coefficient for other disorders, indicating that the patient name item had a closer connection with aphasia than with dysarthria and cognitive disorders. In this way, the *phi* coefficient was obtained for all 29 STAD items.

To estimate item effectiveness, we also calculated the odds ratio (OR); however, we selected the *phi* coefficient for graphing because several OR values reach infinity when zero cell exists. Fisher's exact test was used to analyze whether the presence or absence of a disorder differed depending on the item fail or pass value.

Classification Accuracy

To assess the accuracy of the 3 STAD sections, receiver operating characteristic (ROC) curves were calculated to estimate the optimal cutoff point that provides adequate sensitivity and specificity (patients correctly identified as disorder-positive and disorder-negative, respectively) for screening purposes. ROC curves were generated using the package pROC [58] for R (R Foundation for Statistical Computing, Vienna, Austria) [59]. Dichotomous patient properties (disorder-present or -absent) were used as dependent variables. The total score of the verbal section (range 0–16 [best] points), articulation section (range 0–7), and nonverbal section (range 0–6) were used as independent variables. The area under the ROC curve (AUC) was quantified to determine the overall classification accuracy of each index measure. Conceptually, the AUC equals the probability that a severity score drawn at random from the disorder-present group is higher than that drawn randomly from the disorder-absent group [57]. The performance of the classifier was compared against published benchmarks [60], according to which values >0.9, 0.7–0.9, 0.5–0.7, and 0.5 indicate high accuracy, moderate accuracy, low accuracy, and chance results, respectively.

We used the Youden index (*J*) to identify the optimal cutoff points on the ROC curves [61], defined as follows:

$$J = \max \{ \text{Sensitivity} + \text{Specificity} - 1 \} \quad (2)$$

Especially, $J = 1$ would prove complete separation of the disease-present and -absent groups, while $J = 0$ would indicate complete overlap. *J* represents the value at which $Se + Sp - 1$ is maximized. Once the optimal cutoff points were identified, the positive and negative likelihood ratios were estimated. The former corresponds to the ratio of the probability of correctly or incorrectly classifying a patient as having a disorder (i.e., sensitivity/1 – specificity). The latter corresponds to the ratio of the probability of incorrectly and correctly classifying a patient as disorder-absent (i.e., 1 – sensitivity/specificity) [62].

Correlation with Reference Standard Tests

All correlations between the 3 sections of the STAD and the 3 referenced standard tests were analyzed using Pearson correlation

coefficients to assess the concurrent validity of the STAD and measure whether it is sensitive to disorder severity. The verbal (range 0–16), articulation (range 0–7), and nonverbal (range 0–6) sections were compared with the WAB AQ (range 0–100), AMSD (range 0–29), and WAB NLS (range 0–10) scores, respectively. To test the differences between correlation coefficients, Fisher's *r* to *z* transformation was applied to obtain two-sided probabilities.

R software (version 3.6.2) and Microsoft Excel (Microsoft Corp., Redmond, WA, USA) were used for all statistical analyses. The threshold for statistical significance was set at $p < 0.01$.

Results

Descriptive Statistics

The patients who completed the STAD consisted of 133 women and 181 men with a mean age of 72.7 years (SD = 12.9, range 32–97). The proportion (number) of brain lesions in the left, right, both hemispheres, and cerebellum/brainstem were 42.0% (132), 32.2% (101), 4.1% (13), and 11.5% (36), respectively, while those of atypical lesions and missing data were 8.3% (26), and 1.9% (6), respectively.

Descriptive statistics for the study are presented in Table 5. The table shows statistically significant differences between the disorder-present and disorder-absent groups across the 3 disorders. Group differences were not statistically significant for aphasia and dysarthria with respect to disease duration, defined as the number of days after the onset of brain injury (aphasia: $t(310) = 0.20$, $p = 0.84$, $d = 0.02$; dysarthria: $t(310) = 1.18$, $p = 0.24$, $d = 0.07$), while the difference was significant for cognitive dysfunction ($t(310) = 3.49$, $p < 0.001$, $d = 0.40$). Similarly, the mRS score, which represents the overall stroke severity, with a score of 6 being the worst (dead) and 0 the best (no symptoms), showed statistically significant differences in cognitive dysfunction and dysarthria (aphasia: $t(312) = 2.19$, $p = 0.03$, $d = 0.25$; dysarthria: $t(312) = 3.38$, $p < 0.001$, $d = 0.38$; cognitive dysfunction: $t(312) = 3.38$, $p < 0.001$, $d = 1.06$). These findings suggested that, consistent with other studies [27], the appearance of cognitive dysfunction after stroke could have a negative effect on patients' activities of daily living and result in an extended hospitalization period.

Among all patients, the brain lesions varied according to the disorder (aphasia: $\chi^2(4) = 70.0$, $p < 0.001$, $V = 0.47$; dysarthria: $\chi^2(4) = 14.8$, $p = 0.005$, $V = 0.22$; cognitive dysfunction: $\chi^2(4) = 22.6$, $p < 0.001$, $V = 0.27$). Patients with aphasia were most likely to have a lesion in the left hemisphere of the brain (72 patients vs. an expected frequency of 40.5, obtained from the 2×5 contingency table of presence/ab-

sence of aphasia vs. lesion site). Conversely, the rate of lesions in the right hemisphere was higher in patients with cognitive dysfunction (67 patients vs. an expected frequency of 53.9). Patients with dysarthria had less obvious differences, but the frequency of cerebellar/brainstem lesions exceeded the expected frequency (23 patients vs. 17.5).

When the 3 sections of the STAD were compared in the aphasia group, the differences in all sections were statistically significant. As expected, the effect size of the verbal section was the highest among the 3 sections (verbal section: $t(312) = 15.2, p < 0.001, d = 1.90$; articulation section: $t(312) = 2.1, p = 0.040, d = 0.26$; nonverbal section: $t(312) = 4.4, p < 0.001, d = 0.55$). Among patients in the dysarthria group, the difference was statistically significant in the articulation section, while differences were not significant in the STAD verbal and nonverbal sections (verbal section: $t(312) = 0.1, p = 0.913, d = 0.01$; articulation section: $t(312) = 11.9, p < 0.001, d = 1.36$; nonverbal section: $t(312) = 1.4, p = 0.154, d = 0.16$). Regarding cognitive dysfunction, differences in all sections were statistically significant, and the effect size on the nonverbal section was the highest of the 3 sections (verbal section: $t(312) = 5.7, p < 0.001, d = 0.66$; articulation section: $t(312) = 3.4, p < 0.001, d = 0.39$; nonverbal section: $t(312) = 13.2, p < 0.001, d = 1.51$). There were no differences in age or sex among the 3 disorders.

Item-Level Analysis

Figure 2 shows the effect size obtained by analyzing pass and fail responses on all the 29 STAD items, in 314 patients, to each participant's characteristics of aphasia, dysarthria, and cognitive dysfunction. When comparing the effect size of each item to the 3 disorders, 16, 7, and 6 items belonging to the STAD verbal, articulation, and nonverbal sections, respectively, showed higher values for aphasia, dysarthria, and cognitive dysfunction, respectively. The actual values used in the figure are reported in online supplementary Material S2–S4, which correspond to each of the 3 disorders.

In online supplementary Table A2, all items in the verbal section showed strong correlations with aphasia, and 15 out of 16 items in the STAD verbal section exceeded the moderate effect size of 0.30. The items with a high effect size of 0.5 were the following: naming 2, naming 3, name writing 2, and dictation 1, 2, and 3. Conversely, no item exceeded the effect size of 0.30 on the items belonging to the articulation and nonverbal sections. These results suggested that almost all responses to the verbal section items were related to aphasia, while articulation and nonverbal section items were less related to aphasia.

Fig. 2. Effect size for all Screening Test for Aphasia and Dysarthria (STAD) items among the 3 disorders. The x and y axes of the graph show the ϕ coefficient value and each STAD item, respectively. The orange, green, and navy-blue bars indicate the ϕ coefficients for aphasia, dysarthria, and cognitive dysfunction, respectively.

Regarding dysarthria (online suppl. Table A3), all item responses belonging to the articulation section showed statistically significant differences between the dysarthria-present and dysarthria-absent groups. Four of the 7

Table 6. Classification accuracy of the index test for classifying aphasia, dysarthria, and cognitive dysfunction

STAD	Area under the curve	Optimal cutoff	Sensitivity	Specificity	Positive likelihood ratio ^a	Negative likelihood ratio ^b
Verbal section	0.910	14.5	0.924	0.779	4.19	0.098
Articulation section	0.868	6.5	0.922	0.769	3.99	0.101
Nonverbal section	0.860	5.5	0.821	0.778	3.70	0.230

^a Positive likelihood ratio: sensitivity/(1 – specificity); values of 1 = “neutral,” ≥ 3 = “moderately,” ≥ 10 = “very positive.” ^b Negative likelihood ratio: (1 – sensitivity)/specificity; values of 1 = “neutral,” ≤ 0.30 = “moderately negative,” ≤ 0.10 = “extremely negative” [56].

articulation section items exceeded the effect size of 0.30, while oral diadochokinesis 2 and 3 exceeded the effect size of 0.50. Few items belonging to the verbal and nonverbal sections showed significant differences with respect to the presence or absence of dysarthria, and all of them showed lower effect sizes.

The relationship between cognitive dysfunction and all items in the nonverbal section was significant. Four out of the 6 items of the nonverbal section had an effect size >0.30 , and that of visual construction 2 exceeded 0.5 (online suppl. Table A4). Regarding the relationship between cognitive dysfunction and the other sections, there were significant differences in many items (14 out of 23 total items on verbal and articulation sections corresponded to $p < 0.01$); in particular, dictation 1, 2, and 3 (verbal section) showed moderate effect sizes of 0.31, 0.37, and 0.35, respectively.

Three items (obey command 1, eye contact, and imitation 1), which showed very high correct answer rates, were below the effect size of 0.3 for the targeted disorder, probably because of statistical reasons, as it is well known that skewed distributions have a lower correlation coefficient.

Classification Accuracy

The sensitivity (the rate of disorder-present patients correctly identified as disorder-positive) and specificity (the rate of disorder-absent patients correctly identified as disorder-negative) of the 3 sections of the STAD were assessed for each disorder group. To estimate the classification accuracy of the index test, AUCs were applied to the aphasia-present ($n = 92$) or -absent ($n = 222$) groups for their scores on the STAD verbal section, dysarthria-present ($n = 154$) or absent ($n = 160$) groups for the articulation section, and cognitive dysfunction-present ($n = 179$) or absent ($n = 135$) groups for the nonverbal section. Table 6 provides the AUCs, optimal cutoffs, sensitivity values, specificity values, positive likelihood ra-

Table 7. Pearson’s correlation coefficients between scores on the 3 sections of the index test and the total scores on the WAB AQ, AMSD, and WAB NLS

STAD	WAB AQ ($n = 57$)	AMSD ($n = 84$)	WAB NLS ($n = 71$)
Verbal section	0.891	0.669	0.668
Articulation section	0.456	0.702	0.563
Nonverbal section	0.423	0.658	0.793

Pearson’s correlation coefficient values of 0.30, 0.50, and 0.70 correspond to linear, moderate, and strong linear relationships, respectively. AMSD, Assessment of Motor Speech for Dysarthria; STAD, Screening Test for Aphasia and Dysarthria; WAB AQ, Western Aphasia Battery aphasia quotient; WAB NLS, Western Aphasia Battery nonlinguistic skills.

tios, and negative likelihood ratios for the index measure across the 3 disorders. The AUCs for verbal, articulation, and nonverbal sections were 0.91 (95% CI 0.88–0.94), 0.87 (95% CI 0.81–0.90), and 0.86 (95% CI 0.83–0.90), respectively, indicating moderate to high levels of classification accuracy according to the published benchmarks.

The cutoff scores, which denote the probability thresholds for classifying a patient with brain injury as having one of the 3 disorders, were chosen to optimize sensitivity and specificity for each disorder group at the values of 14.5, 6.5, and 5.5 for the verbal, articulation, and nonverbal sections, respectively. Based on the cutoff point, the sensitivity and specificity values ranged from 0.821 to 0.924 and from 0.769 to 0.779, respectively. Overall, there was a good balance between sensitivity and specificity, with moderate positive and negative likelihood ratios.

Concurrent Validity

We examined the concurrent validity by computing the correlation between the participants’ scores on the 3

STAD sections and 3 other reference tests, which were conducted by licensed SLPs within 2 weeks after STAD administration. Of the 212 patients who underwent the reference test, 57, 84, and 71 patients were administered WAB Parts I–IV to calculate AQ, AMSD, and WAB Part VIII for calculating NLS, respectively. The means, SD, and score ranges from patients with brain injury to the reference measures are provided in online supplementary Material S5.

Table 7 presents the Pearson's correlation coefficients between all the scores of these measures. Although all correlations showed significance at $p < 0.01$, the r coefficients differed depending on the combination of the STAD section and the reference test. When the STAD verbal section score was correlated with the WAB AQ, AMSD, and WAB NLS, the strongest correlation was that with the WAB AQ (WAB AQ: $r = 0.89$, 95% CI 0.82–0.93; AMSD: $r = 0.67$, 95% CI 0.53–0.77; WAB NLS: $r = 0.67$, 95% CI 0.51–0.78), while in the case of the articulation section score, the correlation with AMSD was the strongest among all combinations (WAB AQ: $r = 0.46$, 95% CI 0.22–0.64; AMSD: $r = 0.70$, 95% CI 0.57–0.80; WAB NLS: $r = 0.56$, 95% CI 0.38–0.70). In the case of the nonverbal section, the correlation with WAB NLS was the strongest (WAB AQ: $r = 0.42$, 95% CI 0.18–0.62; AMSD: $r = 0.66$, 95% CI 0.52–0.76; WAB NLS: $r = 0.79$, 95% CI 0.69–0.87). In our tests of differences between the correlation coefficients, the difference between the scores on the STAD verbal test and WABAQ was greater than that between the scores on the AMSD and WABNS ($p < 0.05$), the difference between the articulatory test and the AMSD was greater than that with the WABAQ ($p < 0.05$), and the difference between the nonverbal test and the WABNS was greater than that with the WABAQ ($p < 0.05$). No change was observed in the concurrent validity due to the time after onset (data not shown).

Discussion

The STAD takes multiple domains into account, specifically language, articulation, and cognitive function, because all these evaluations are frequently required to assess communicative function after brain injury [63, 64]. This study examined the hypotheses that, in terms of aphasia, dysarthria, and cognitive disorder, the STAD could show acceptable validity. We confirmed the hypotheses with regard to item-level validity, diagnostic accuracy, and concurrent validity using 314 STAD and 212 other reference test data from 20 centers across Japan.

We applied the *phi* coefficient to item-level analysis, as it showed the effectiveness of each item with regard to each type of communicative disorder: the higher the coefficient value, the closer the relationship between the item and the disorder. The *phi* coefficients of 23 out of 29 items exceeded the moderate effect size of 0.3 toward the targeted disorder. Items with high effect sizes of 0.5, such as naming 2, naming 3, name writing 2, and dictation 1, 2, and 3 for aphasia, oral diadochokinesis 2 and 3 for dysarthria, and visual construction 2 for cognitive disorder, were considered good items for evaluating those disorders. Cohen's *d* values (which express how distant the test performance is between those with and without the disorder), exceeded 0.8 (large effect), on the STAD's verbal, articulation, and nonverbal sections, to evaluate aphasia, dysarthria, and cognitive disorder, respectively. The accuracies for differentiating patients with and without disorders based on the 3 STAD sections were moderate to high across our results (concerning the AUCs, sensitivity and specificity values, and positive/negative likelihood ratios). When comparing the STAD to the reference tests for aphasia, dysarthria, and cognitive dysfunction evaluation (WAB AQ, AMSD, and WAB NLS, respectively), we found strong correlations between the STAD verbal, dysarthria, and nonverbal sections and the corresponding tests, which demonstrated good criteria-related validity.

El Hachoui et al. [20] evaluated the degree of bias in validity studies based on 3 aspects: blinding, consecutiveness, and representativeness. In our study, the same examiner conducted the clinical diagnosis and the reference test. Therefore, blinding was insufficient. Our study included consecutive cases of brain injury, thus yielding correct consecutiveness, which is essential to reduce selection bias in screening validity studies. According to Hachoui's review, among past studies on communicative screening, Kostálová et al. [65] used the largest cohort size (i.e., 149 cases), which was exceeded by our sample size (i.e., 316 cases). Further, we clearly defined the demographic and clinical profiles as well as the general clinical and functional characteristics of the patients. Therefore, this study has adequate representativeness. Based on these factors, the risk of bias in our study can be regarded as intermediate. Additionally, although the evaluation method proposed by El Hachoui et al. [20] does not include single- or multiple-facility study designs, the multicenter nature of our study reduced the bias deriving from population differences among clinical environments or geographic factors, thus enhancing the reproducibility of the performance of the STAD and that of the reference standard tests.

The presence of 3 test sections in the STAD that respond specifically to 3 different types of communication disorders may alleviate the problem of false positives observed in existing screening tests [23, 29, 35–38]. Additionally, using the results of independent sections makes it easy to estimate the most important factors that inhibit patient communication. For example, the false positives seen in the STAD verbal and articulation tests due to cognitive dysfunction are unavoidable (in particular, dictation 1, 2, and 3 showed a moderate effect size for cognitive dysfunction). However, the STAD includes a nonverbal test that strongly corresponds to cognitive function. Therefore, when the nonverbal test scores are significantly lower than those of the language/articulation test, we can assume that cognitive dysfunction is the most significant factor impeding communication, which makes it easier to detect false positives that may occur in language and articulation tests. One advantage of having a screening test for multiple areas is that it can deal with a diverse range of communication alterations. The STAD makes it easier to assess these comorbid disorders.

A similar test, also dealing with a diverse range of communication disorders, is the SCCAN, which can also be administered at the bedside, in an exam room, or at home [34, 66]. The purpose and composition of this test are similar to those of the STAD. The SCCAN is already marketed in Western countries and has been introduced in the Mental Measurements Yearbook as a test with sufficient reliability and validity [67]. However, the SCCAN does not attempt to address motor-speech function. Moreover, the time required to complete the SCCAN is much longer than that required to complete the STAD; the SCCAN requires approximately 35 min to complete, as it comprises of 94 items organized in 8 scales; in contrast, the STAD, composed of 29 items organized in 3 sections, could be completed in 9 min 48 s on average [39]. Therefore, the STAD would be suitable for less robust patients, including those in hyperacute environments (e.g., intensive care unit), those with more severe conditions, and those with a tendency toward psychological instability. Conversely, the SCCAN provided more accurate test results. According to a previous study, the sensitivity and specificity rates for the SCCAN were 98 and 95%, respectively, when comparing 40 healthy controls with 51 patients (20, 15, and 16 with left hemisphere pathology, right hemisphere pathology, and Alzheimer's disease, respectively) [66]. There is a tradeoff in the development of tests, with shorter tests being less demanding on the patient, and longer tests being preferable when the goal is to

obtain more information. Each therapist should select the appropriate test after ascertaining the patient's environment and condition.

Our study had 3 main limitations. First, although the International Council for Harmonization Guidelines on Statistical Principles for Clinical Trials [68] emphasize that it is especially important for rating scales to address content validity and inter- and intra-rater reliability, responsiveness to changes in disease severity and reliability have not yet been reported for the STAD. Therefore, we plan to conduct video-based inter-rater reliability testing. Second, although the AMSD is not used worldwide, we used it as a reference standard test because it is the only standard test available in Japan for assessing dysarthria. Further studies are required to assess whether the STAD is correlated with a global, gold-standard articulation test. Third, we recognized that a comparison of the SCCAN with other tests should have been discussed at the beginning of this study; however, since a Japanese version of the SCCAN has yet to be developed, we could not perform any experiments to address it. In the future, it is imperative to test both the STAD and SCCAN in a single population and compare the results to expand the understanding of the features of these tests and serve the frequent requirement of observing multiple communicative functions in cases of brain injury.

In conclusion, this short screening tool can be very useful in specific contexts, such as in early bedside investigations, to obtain a quick summarized assessment of communicative function prior to the administration of the other tests, and in cases where more in-depth testing is not feasible. By evaluating a wider range of neurological communicative disorders in patients with brain injury, the STAD could be a sensitive tool for these settings and populations.

Acknowledgment

We would like to thank the following individuals for their helpful consultation role during the STAD development: Akatsuki Sakaide, Akihito Morimune, Akiko Tachibana, Akiko Takahashi, Akira Yokota, Asuka Mizuno, Chie Inamori, Chika Teramoto, Chikako Okano, Daisuke Furukawa, Hiroaki Yamada, Hironao Ehara, Hiroshi Yamamoto, Hiroto Watanabe, Hisakazu Sato, Hounari Takayanagi, Kayoko Mori, Keita Takenaka, Kazuhiko Asada, Kiwamu Teruya, Kouhei Tsubota, Kumiko Yamamoto, Madoka Imagawa, Maiko Tanaka, Makoto Nagasawa, Manabu Abe, Masaki Tahahashi, Masashi Yamaguchi, Miho Miyasaka, Naoki Ishibashi, Naomi Okamura, Noriyuki Komori, Reiko Eguchi, Reiko Kimura, Rui Nakajima, Sachiyo Muranishi, Satoru Tanabe,

Satsuki Hashimoto, Shoko Marui, Shota Nishioka, Sonoko Uno, Syoko Nitou, Syota Nishioka, Takeshi Hukuma, Tatsuya Nemoto, Tomoki Hamada, Tomoko Hukunaga, Tomomi Santa, Tsuyoshi Nakajima, Wakana Nakano, Yasuyuki Mori, Yoshihumi Higashi, Yosuke Kurokawa, Yudai Sakai, Yuka Murakami, Yuriko Ishibasi, and Yuya Hashimoto.

Statement of Ethics

This study adhered to the Declaration of Helsinki guidelines for experiments involving humans. The Ethics Review Committee of Chiba University approved this study (ID: 2627). Written informed consent was obtained from all the participants.

Conflict of Interest Statement

K.A. and M.K. received royalties from the publication of the Japanese version of the STAD. The remaining authors have no conflicts of interest to declare.

Funding Sources

This study was supported by the JSPS Kakenhi, grant No. 16H00709, 19K24192, and Pfizer Academic Contribution. The funding bodies had no role in the conduct or reporting of the research.

Author Contributions

K.A. and Y.H. performed the analysis and interpreted the data. K.A., M.K., and J.F. conceived and developed the STAD and planned the study. E.S. contributed to the conceptualization and supervision of the project. K.A. took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript. All authors approved the final version of the manuscript.

Data Availability Statement

The data that support the findings of this study are available at <https://doi.org/10.6084/m9.figshare.14898774>, reference number [69].

References

- Papathanasiou I, Coppens P. Aphasia and related neurogenic communication disorders. 2nd ed. Burlington, MA; Jones & Bartlett learning; 2016.
- Engelter ST, Gostynski M, Papa S, Frei M, Born C, Ajdacic-Gross V, et al. Epidemiology of aphasia attributable to first ischemic stroke: incidence, severity, fluency, etiology, and thrombolysis. *Stroke*. 2006 Jun;37(6):1379–84.
- Laska AC, Hellblom A, Murray V, Kahan T, Von Arbin M. Aphasia in acute stroke and relation to outcome. *J Intern Med*. 2001 May;249(5):413–22.
- Bhatnagar SC, Andy OJ. Neuroscience for the study of communicative disorders. Baltimore: Williams & Wilkins; 1995.
- Duffy JR. Motor speech disorders e-book. 4th ed. Elsevier Health Sciences; 2019. p. 3.
- Lawrence ES, Coshall C, Dundas R, Stewart J, Rudd AG, Howard R, et al. Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke*. 2001 Jun;32(6):1279–84.
- Warlow C, van Gijn J, Dennis M, Wardlaw J, Bamford J, Hankey G, et al. Stroke: Practical management. 3rd ed. Oxford: Blackwell; 2008. <https://doi.org/10.1002/9780470696361>
- Wang YT, Kent RD, Duffy JR, Thomas JE. Dysarthria associated with traumatic brain injury: speaking rate and emphatic stress. *J Commun Disord*. 2005 May-Jun;38(3):231–60.
- Urban PP, Rolke R, Wicht S, Keilmann A, Stoeter P, Hopf HC, et al. Left-hemispheric dominance for articulation: a prospective study on acute ischaemic dysarthria at different localizations. *Brain*. 2006 Mar;129(Pt 3):767–77.
- Patel M, Coshall C, Rudd AG, Wolfe CD. Natural history of cognitive impairment after stroke and factors associated with its recovery. *Clin Rehabil*. 2003 Mar;17(2):158–66.
- Nys GM, Van Zandvoort MJ, De Kort PL, Jansen BP, Van der Worp HB, Kappelle LJ, et al. Domain-specific cognitive recovery after first-ever stroke: a follow-up study of 111 cases. *J Int Neuropsychol Soc*. 2005 Nov;11(7):795–806.
- Riepe MW, Riss S, Bittner D, Huber R. Screening for cognitive impairment in patients with acute stroke. *Dement Geriatr Cogn Disord*. 2004;17(1-2):49–53.
- Kimbarow ML. Cognitive communication disorders. 3rd ed. San Diego: Plural Publishing Inc; 2019.
- Nouwens F, Visch-Brink EG, Van de Sandt-Koenderman MM, Dippel DW, Koudstaal PJ, de Lau LM. Optimal timing of speech and language therapy for aphasia after stroke: more evidence needed. *Expert Rev Neurother*. 2015;15(8):885–93.
- Mackay LE, Bernstein BA, Chapman PE, Morgan AS, Milazzo LS. Early intervention in severe head injury: long-term benefits of a formalized program. *Arch Phys Med Rehabil*. 1992 Jul;73(7):635–41.
- Cope DN, Hall K. Head injury rehabilitation: benefit of early intervention. *Arch Phys Med Rehabil*. 1982 Sep;63(9):433–7.
- Teasell R, Bitensky J, Salter K, Bayona NA. The role of timing and intensity of rehabilitation therapies. *Top Stroke Rehabil*. 2005;12(3):46–57.
- Geis C, Grissom S. Rehabilitation Medicine Issues in Acute Stroke: prevention of common complications, facilitation of recovery, and assessment of rehabilitation potential. *Neurologist*. 1997;3(5):322.
- Doesborgh SJ, van de Sandt-Koenderman WM, Dippel DW, van Harskamp F, Koudstaal PJ, Visch-Brink EG. Linguistic deficits in the acute phase of stroke. *J Neurol*. 2003 Aug;250(8):977–82.
- El Hachoui H, Visch-Brink EG, de Lau LM, van de Sandt-Koenderman MW, Nouwens F, Koudstaal PJ, et al. Screening tests for aphasia in patients with stroke: a systematic review. *J Neurol*. 2017 Feb;264(2):211–20.
- Ross KB, Wertz RT. Discriminative validity of selected measures for differentiating normal from aphasic performance. *Am J Speech Lang Pathol*. 2003 Aug;12(3):312–9.
- Hénon H, Lebert F, Durieu I, Godefroy O, Lucas C, Pasquier F, et al. Confusional state in stroke: relation to preexisting dementia, patient characteristics, and outcome. *Stroke*. 1999 Apr;30(4):773–9.
- Salter K, Jutai J, Foley N, Hellings C, Teasell R. Identification of aphasia post stroke: a review of screening assessment tools. *Brain Inj*. 2006 Jun;20(6):559–68.
- Crary MA, Haak NJ, Malinsky AE. Preliminary psychometric evaluation of an acute aphasia screening protocol. *Aphasiology*. 1989 Mar;3(7):611–8.

- 25 Enderby P, Wood VA, Wade DT, Hewer RL. Aphasia after stroke: a detailed study of recovery in the first 3 months. *Int Rehabil Med*. 1987;8(4):162–5.
- 26 NHS [Internet]. Management of patients with stroke: Rehabilitation, prevention and management of complications, and discharge planning Scottish Intercollegiate Guidelines Network Part of NHS Quality Improvement Scotland. 2010 [cited 2020 Dec 11]. Available from: www.sign.ac.uk
- 27 Hinckley JJ. A case for the implementation of cognitive-communication screenings in acute stroke. *Am J Speech Lang Pathol*. 2014 Feb;23(1):4–14.
- 28 Malloy PF, Cummings JL, Coffey CE, Duffy J, Fink M, Lauterbach EC, et al. Cognitive screening instruments in neuropsychiatry: A report of the Committee on Research of the American Neuropsychiatric Association. *J Neuropsychiatry Clin Neurosci*. 1997;9(2):189–97.
- 29 Nakase-Thompson R, Manning E, Sherer M, Yablon SA, Gontkovsky SL, Vickery C. Brief assessment of severe language impairments: initial validation of the Mississippi aphasia screening test. *Brain Inj*. 2005 Aug;19(9):685–91.
- 30 El Hachoui H, Sandt-Koenderman MW, Dippel DW, Koudstaal PJ, Visch-Brink EG. The ScreeLing: occurrence of linguistic deficits in acute aphasia post-stroke. *J Rehabil Med*. 2012 May;44(5):429–35.
- 31 Enderby PM, Wood VA, Wade DT, Hewer RL. The Frenchay Aphasia Screening Test: a short, simple test for aphasia appropriate for non-specialists. *Int Rehabil Med*. 1987;8(4):166–70.
- 32 Management of Stroke Rehabilitation Working Group. VA/DoD clinical practice guideline for the management of stroke rehabilitation. version 2.0. Washington (DC): Veterans Health Administration and Department of Defense; 2010.
- 33 Lindsay MP, Gubitz G, Bayley M, Hill MD, Davies-Schinkel C, Singh S, et al. Canadian best practice recommendations for stroke care (update 2010). On behalf of the Canadian Stroke Strategy Best Practices and Standards Writing Group. Ottawa, Ontario, Canada: Canadian Stroke Network; 2010.
- 34 Milman LH, Holland A. Scales of cognitive and communicative ability for neurorehabilitation. Dallas: PRO-ED, Inc; 2012.
- 35 Thommessen B, Thoresen GE, Bautz-Holter E, Laake K. Screening by nurses for aphasia in stroke—the Ullevaal Aphasia Screening (UAS) test. *Disabil Rehabil*. 1999 Mar;21(3):110–5.
- 36 Al-Khawaja I, Wade DT, Collin CF. Bedside screening for aphasia: a comparison of two methods. *J Neurol*. 1996 Feb;243(2):201–4.
- 37 Gaber TA, Parsons F, Gautam V. Validation of the language component of the Addenbrooke's Cognitive Examination—Revised (ACE-R) as a screening tool for aphasia in stroke patients. *Australas J Ageing*. 2011 Sep;30(3):156–8.
- 38 Gibson L, MacLennan WJ, Gray C, Pentland B. Evaluation of a comprehensive assessment battery for stroke patients. *Int J Rehabil Res*. 1991;14(2):93–100.
- 39 Araki K, Kozono M. Screening test for aphasia and dysarthria. Interuna Publishers Inc; 2018. [(in Japanese)].
- 40 Blumenfeld H. Neuroanatomy through clinical cases. Neuroanatomy through clinical cases. 2nd ed. Sunderland (MA): Sinauer Associates Inc; 2010. pp. 830–1.
- 41 Campbell WW, Title N (Williams W, Campbell MD, editors). DeJong's the neurologic examination. Philadelphia (PA): Lippincott; 2006. p. 79.
- 42 Araki K, Uno S, Fujitani J, Fushimi T. A screening test for aphasia and dysarthria for patients with acute stroke. *Nihon Gengo Chōkakushi Kyōkai (Japanese J Speech, lang hear res)*. 2009;6:3–11.
- 43 Araki K, Kozono M, Asada K, Shimizu E. Standardization of a Screening Test for Aphasia and Dysarthria (STAD). *Clin Neuropsychol*. 2019;29:11–6.
- 44 Itaya T, Murakami Y, Ota A, Nomura E, Fukushima T, Nishigaki M. Assessment model to identify patients with stroke with a high possibility of discharge to home: a retrospective cohort study. *Stroke*. 2017 Oct;48(10):2812–8.
- 45 Sugishita M. Western aphasia battery (WAB). Tokyo: Igaku-Shoin; 1986. [(in Japanese)].
- 46 Kertesz A. Western aphasia battery test manual. New York: Grune & Stratton; 1982.
- 47 Nishio M. Assessment of motor speech disorders (AMSD). Tokyo: Interuna Publishers Inc; 2004. [(in Japanese)].
- 48 Hasegawa T, Kishi H, Shigeno K, Tanemura J, Kusunoki T. Three-dimensional structure in language tests of aphasia. *Folia Phoniatr (Basel)*. 1985;37(5-6):246–58.
- 49 De Renzi E, Vignolo LA. The token test: A sensitive test to detect receptive disturbances in aphasics. *Brain*. 1962 Dec;85(4):665–78.
- 50 Mori E, Mitani Y, Yamadori A. Usefulness of a Japanese version of the Mini-Mental State Test in neurological patients [(in Japanese)]. *Jpn J Neuropsychol*. 1985;1:82–90.
- 51 Kohs SC. The Block-Design Tests. *J Exp Psychol*. 1920;3(5):357–76.
- 52 Raven JC, Court JH, Raven JC. Raven's progressive matrices and vocabulary scales. Oxford: Oxford Psychologists Press; 2008. pp. 1938–98.
- 53 Wilson BA, Cockburn J, Baddeley A. Rivermead Behavioural Memory Test. London: Thames Valley Test Company; 1985.
- 54 Reitan RM. The relation of the trail making test to organic brain damage. *J Consult Psychol*. 1955 Oct;19(5):393–4.
- 55 Wechsler D. Wechsler Adult Intelligence Scale. 3rd ed. Wechsler Psychological Corp; 1997.
- 56 Grant DA, Berg EA. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J Exp Psychol*. 1948 Aug;38(4):404–11.
- 57 Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r. *Law Hum Behav*. 2005 Oct;29(5):615–20.
- 58 Robin X, Turck N, Hainard A, Tiberti N, Lissacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011 Mar;12(1):77.
- 59 R Development Core Team [Internet]. R A language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing. - References [cited 2020 Dec 10]. Available from: [https://www.scrip.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/referencespapers.aspx?referenceid=2600003](https://www.scrip.org/(S(351jmbntvnsjt1aadkposzje))/reference/referencespapers.aspx?referenceid=2600003). Scientific Research Publishing.
- 60 Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr*. 2007 May;96(5):644–7.
- 61 Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950 Jan;3(1):32–5.
- 62 McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002 Aug;17(8):646–9.
- 63 Coelho CA, DeRuyter F, Stein M. Treatment efficacy: cognitive-communicative disorders resulting from traumatic brain injury in adults. *J Speech Hear Res*. 1996 Oct;39(5):S5–17.
- 64 Shipley KG, McAfee JG. Assessment of neurologically based communicative disorders. Assessment in speech-language pathology: A resource manual. New York: Delmar Publishers; 2008. p. 407.
- 65 Kostálová M, Bártková E, Sajgalíková K, Dolenská A, Dusek L, Bednarík J. A standardization study of the Czech version of the Mississippi Aphasia Screening Test (MASTcz) in stroke patients and control subjects. *Brain Inj*. 2008 Sep;22(10):793–801.
- 66 Milman LH, Holland A, Kaszniak AW, D'Agostino J, Garrett M, Rapcsak S. Initial validity and reliability of the SCCAN: using tailored testing to assess adult cognition and communication. *J Speech Lang Hear Res*. 2008 Feb;51(1):49–69.
- 67 Carlson JF, Geisinger KF, Jonson JL, Title N. Scales of cognitive and communicative ability for neurorehabilitation. The twentieth mental measurements yearbook. Lincoln (Nebr): Burros Institute of Mental; 2017. pp. 634–8.
- 68 Guidelines IC. General considerations for clinical trials E8. In international conference on harmonization of technical requirements for registration of pharmaceuticals for human use [1997]. Available at: <https://www.ema.europa.eu/en/ich-e9-statistical-principles-clinical-trials>.
- 69 Araki K, Hirano Y, Kozono M, Fujitani J, Shimizu E. Dataset of the Screening Test for Aphasia and Dysarthria (STAD): consecutively selected brain injury patients from 20 medical centers in Japan. 2021. <http://doi.org/https://doi.org/10.6084/m9.figshare.14898774>