

ALSWAITI, FADI Y., Ph.D. Content Validation of the JAAT: An Aphasia Assessment Designed for Jordanian-Arabic Speaking Population. (2019)
Directed by Dr. Robert Mayo. 154 pp.

Aphasia has significant consequences on communication in activities of daily living and comprehensive aphasia assessments are an important clinical tool for gathering diagnostic and treatment information. A comprehensive aphasia assessment is still the most used tool utilized when the aim is to collect as much clinically useful information as possible using one single test. Despite the availability of several major comprehensive aphasia tests for the English-speaking population, there are only a few comprehensive tests that have been developed for the majority of the world's popular languages including Arabic (Ivanova & Hallowell, 2013). Clinicians in Jordan are challenged by the lack of availability of a formal aphasia test for people with aphasia (PWA) speaking a Jordanian dialect of Arabic. The aim of this study was to examine the content validity of a newly developed Jordanian-Arabic Aphasia Test (JAAT) by formulating an experimental process of content validation. The study consisted of three experiments: the first experiment aims to validate the visual stimuli used in the JAAT; the second to identify troublesome in need of modification; and the third is an examination of content validity of the JAAT.

The JAAT consists of 12 subtests divided into four parts with a total of 112 items that assess verbal fluency, auditory comprehension, repetition, and naming. For the first experiment, 164 non-brain damaged (NBD) Jordanian-Arabic speakers distributed by age, sex, and level of education were asked to match 132 pictures with target words to determine whether the images accurately represent the words they were drawn to depict.

In the second experiment the JAAT was administered to 20 NBD Jordanian Arabic speakers who were 60 years of age or more to identify item outliers that are less than ideal for use with PWA. As for the third experiment and the focus of this study, 21 subject matter experts (SME) in the field of speech-language pathology were asked to rate the JAAT using an online questionnaire. The content of the JAAT was appraised using three parameters—clinical applicability, linguistic relevance, and cultural sensitivity. Furthermore, a Content Validity Index (CVI) with a criterion set for 80% was used based on ratings of 7 or 6 indicating response choices of strongly agree and agree respectively on the online questionnaire.

Results of the first experiment demonstrated that the majority of the items in the JAAT were validated by NBD Jordanian Arabic speakers and found to accurately represent the target words. In the second experiment, as expected, senior Jordanian-Arabic speakers performed correctly on most items of the JAAT; with 8 items shown to be difficult for more than one-fourth of examinees. These items were modified before being readmitted in the JAAT for content validation. Results of the third experiment showed that the SMEs rated the JAAT's clinical applicability, linguistic relevance, and cultural sensitivity positively with significant CVI values. Two subtests in the JAAT were found to require further modifications and the test's capacity to identify other disorders was found to be lacking. Additionally, the JAAT needs to go through some modification to improve its potential to identify severity levels of PWA. The results of this study signal the importance of implementing a scheme of content validation and item development process that paves the way for better test construction practices.

CONTENT VALIDATION OF THE JAAT: AN APHASIA ASSESSMENT DESIGNED
FOR JORDANIAN-ARABIC SPEAKING POPULATION

by

Fadi Y. AlSwaiti

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2019

Approved by

Committee Chair

To Hana and our kids,
to both my mother and my mother-in-law,
and to my sister Jehan

APPROVAL PAGE

This dissertation, written by Fadi Y. AlSwaiti, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Robert Mayo

Committee Members _____
Mark Schulz

Denise Tucker

Mary Compton

Date of Acceptance by Committee

Date of Final Oral Examination

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
I. STATEMENT OF THE PROBLEM	1
II. REVIEW OF LITERATURE	6
Aphasia	6
Definition	6
Subgroups	7
Epidemiology	8
Assessments Used to Assess Aphasia	10
Screening Tests	12
Functional Aphasia Tests	12
Comprehensive Aphasia Tests	13
Translating and Adapting English Tests	15
Arabic Aphasia Tests	16
The Arabic Diagnostic Aphasia Battery (A-DAB)	17
The Assessment of Language-related Functional Activities (ALFA)	18
The Object and Action Naming Battery (OANB)	19
The Bilingual Aphasia Battery (BAT)	19
The Egyptian-Arabic Comprehensive Aphasia Test (E-CAT)	20
The Qatari-Arabic Comprehensive Aphasia Test (Q-CAT)	21
Content Validity	22
The Jordanian Aphasia Assessment Test (JAAT)	24
Elements of the JAAT	27
Statement of Purpose	34
III. METHOD	38
Institutional Approval	38
Study Design	39

Experiment One (Questionnaire 1: Verification of Visual Stimuli of the JAAT-1 by NBD Jordanian Arabic Speakers)	40
Participants.....	40
Measures	42
Procedures.....	42
Data Analysis for Experiment One Questionnaire.....	42
Experiment Two (Identifying Inadequate Items in the JAAT-2 by Administering the Test on Senior NBD Jordanian Arabic Speakers)	43
Participants.....	43
Measures	44
Procedures for Experiment Two	44
Scoring of the JAAT-2 and Data Analysis of Experiment Two	45
Experiment Three (Questionnaire 2: Validation of Test Content Based on SMEs' Ratings of the JAAT-3)	47
Participants.....	47
Measures for Experiment Three.....	48
Procedures for Experiment Three	49
Data Analysis for Experiment Three	50
 IV. RESULTS	 53
Experiment One (Questionnaire 1: Verification of the JAAT-1 Visual Stimuli by NBD Jordanian Arabic Speakers)	53
Experiment Two (Identifying Inadequate Items in the JAAT-2 by Administering the Test on Senior NBD Jordanian Arabic Speakers)	54
Experiment Three (Questionnaire 2: Validation of Test Content Based on SME Rating of the JAAT-3).....	59
Series One Statistical Analysis: Examining the Clinical Applicability of the JAAT-3	60
Series Two Statistical Analysis: Examining Linguistic and Cultural Sensitivity and Overall Impressions of the JAAT-3.....	69
Series Three Statistical Analysis: Comparing the Performance of the Four Parts of the JAAT-3 to Each Other.....	77
Series Four Statistical Analysis: Examining the Consistency of SME Responses on Questionnaire 2	79

V. DISCUSSION	81
Clinical Implications	87
Limitations and Future Directions	89
Summary and Conclusions	89
REFERENCES	92
APPENDIX A. THE JORDANIAN-ARABIC APHASIA TEST (JAAT).....	113
APPENDIX B. SAMPLE ITEM IN QUESTIONNAIRE 1 OF EXPERIMENT ONE	144
APPENDIX C. QUESTIONNAIRE 2 ITEMS APPRAISING THE JAAT AS A UNIT	145
APPENDIX D. SAMPLE OF SUBTEST APPRAISING QUESTIONNAIRE ITEMS	146
APPENDIX E. SAMPLE OF PART APPRAISING QUESTIONNAIRE ITEM.....	148
APPENDIX F. QUESTIONNAIRE ITEMS APPRAISING THE JAAT AS A UNIT	149
APPENDIX G. MODIFIED OR ADDED ITEMS TO JAAT AFTER EXPERIMENT TWO.....	152

LIST OF TABLES

	Page
Table 1. Subtest Points and Scoring Methods	46
Table 2. Number of Words Distributed by Percentages of Responses on Questionnaire 1	54
Table 3. Average Performances of Subjects on the JAAT-2.....	55
Table 4. Items with 30% or More Incorrect Responses on the JAAT-2	58
Table 5. Questionnaire Items Appraising Clinical Applicability of Subtests to Test Aphasia in Jordanian Arabic Speakers	61
Table 6. Results Showing Agreement as to Whether Each Part Contains Enough Items in its Subtest to Assess What it Was Designed to Assess	65
Table 7. Questionnaire Items Appraising the JAAT-3 as a Testing Unit.....	66
Table 8. Questionnaire Items Appraising the JAAT-3 as a Testing Unit.....	68
Table 9. Questionnaire Items Appraising Important Clinical Aspect of the JAAT-3 as a Unit.....	68
Table 10. Questionnaire Items Appraising “Linguistic Relevance” of Subtests to Jordanian Arabic Speakers	70
Table 11. Questionnaire Items Appraising “Cultural Sensitivity” of Subtests to Jordanian Arabic Speakers	73
Table 12. Questionnaire Items Appraising the “Overall” Impression of SME on Subtests.....	75
Table 13. Reliability of Questionnaire Items	80

LIST OF FIGURES

	Page
Figure 1. Data Distribution of the 21 SMEs Responses on Clinical Applicability of Subtests in the JAAT-3.....	63
Figure 2. Range Distribution of Response Scores on Items Assessing Clinical Attributes of the JAAT-3	69
Figure 3. Data Distribution of the 21 SMEs Responses on Linguistic Relevance of Subtests in the JAAT-3.....	71
Figure 4. Data Distribution of the 21 SMEs Responses on Cultural Sensitivity of Subtests in the JAAT-3	74
Figure 5. Data Distribution of the 21 SMEs Responses on Overall Impressions of Subtests in the JAAT-3	76
Figure 6. Mean Response Scores on Subtests within the Four Parts of the JAAT-3.....	78
Figure 7. Mean Response Subtests' Scores of Parts with Percentile Agreement on Parts Questions.....	79

CHAPTER I

STATEMENT OF THE PROBLEM

Arabic language contains a variety of spoken dialects and a standard written language (Bikel, 2002). The dialects show phonological, morphological, lexical, and syntactic differences that can be slightly comparable to those among the Romance languages in Europe (Chiang, Diab, Habash, Rambow, & Shareef, 2006). The standard written language, Modern Standard Arabic (MSA), is the same throughout the Arab world and used in some scripted communication in formal settings such as newscasting and political debates (Chiang et al., 2006). In the past several years, there has been a growing interest in developing diagnostic Arabic tests of language disorders. There are now several Arabic diagnostic tests for assessing child and adult language disorders in Arabic populations. Egypt has led the way in this development. In 2009 Abo Ras, Aref, El-Raghy, Gaber, and El-Maghraby developed the Comprehensive Test of Arabic Language as a tool to determine language delay in Egyptian children. As for adults with aphasia, two diagnostic tests for adults with aphasia are currently available in Egypt, the Kasr El-Aini Arabic Aphasia Test (KAAT; Hassanein et al., 2002) and an Egyptian-Arabic adaptation of the Comprehensive Aphasia Test (CAT; Swinburn, Porter, & Howard, 2005; Abou El-Ella et al., 2013).

More recently, a bedside screening tool named the Arabic Diagnostic Aphasia Battery (A-DAB-1; Al-Thalaya et al., 2018) was developed for the Lebanese population.

The Gulf region saw the development of three tests: A translation of the Assessment of Language-Related Functional Activities (ALFA; Baines, Heeringa, & Martin, 1999; Al Yaari & Almaflehi, 2013); the Aphasia Battery for Qatari-Arabic (Khwaileh, Mustafawi, Howard, & Herbert, 2016); and an adapted version of the Object and Action Naming Battery for the Saudi population (OANB; Alyahya & Druks, 2016; Druks & Masterson, 2000). Although Jordan was the first to adapt a test for aphasia with the Bilingual Aphasia Test (BAT; Paradis & Abidi, 1987; Paradis & El Halees, 1989), the test was never normed on Arabic speakers, with or without aphasia (Ivanova & Hallowell, 2013) and it is currently not available for clinicians.

Currently, no comprehensive aphasia test is available in Jordan, so the need for a Jordanian Arabic Test of aphasia is pressing. A valid and comprehensive language assessment is a vital expectation in post-stroke rehabilitation for individuals with aphasia (Gialanella, 2011; Hersh, Wood, & Armstrong, 2018; Heuer, Ivanova, & Hallowell, 2017). Considering that one of the most important requirements for tests of any type is standardization (Gregory, 2007), the lack of standardized tests for people with aphasia (PWA) in Jordan leads speech-language pathologists (SLPs) to be highly dependent on non-standardized testing methods. These methods do not provide the subtle and formal distinctions that standardized tests provide between normal and disordered performances (Spreen & Risser, 2003). Another practice that clinicians may be forced to use due to the lack of standardized tests is informal translations of standardized aphasia tests (Khoja, 2017). Using tools that were designed and normed for other cultural and linguistic

populations might lead to results that do not accurately reflect the performances of examinees.

Over the past several years, with input from Jordanian SLPs and colleagues, four Subject Matter Experts (SMEs) and with design influence from the third version of the Boston Diagnostic Aphasia Examination (BDAE-3; Goodglass, Kaplan, & Baresi, 2001), this author has designed an aphasia test in the Jordanian Arabic dialect named the Jordanian Arabic Aphasia Test (JAAT). The early version was designed to be used with both the Saudi and Jordanian populations. In the past five years this test has been modified so it can be more accommodating to the Jordanian population. The JAAT consists of four parts: Verbal fluency, naming, repetition, and auditory comprehension. There are 112 items divided into 12 subtests comprising the four parts. The complete test appears in Appendix A.

The purpose of this present study was to examine the content validity of the JAAT by conducting an experimental process of content validation. To achieve this the study was divided into three experiments. The first experiment attempted to validate 132 pictures used or those that potentially can be used in the JAAT by asking 164 non-brain damaged (NBD) Arabic-Jordanian speakers to match the pictures with the words they represented. This was achieved by an online questionnaire and pictures that did not achieve 90% agreement were removed or modified. The second experiment sought to identify test items that prove to be too difficult to 30% or more of the 20 NBD subjects 60 years or more in age. The JAAT was administered to the NBD subjects and items that proved to be inappropriate for use with PWA due to their high difficulty were either

removed or modified. The third experiment examined content validity of all subtests and parts of the JAAT along with the JAAT as a unified testing unit. The third experiment had 21 subject matter experts (SME) in the field of speech-language pathology in Jordan using an online questionnaire to rate the JAAT. The content of the JAAT was assessed by SMEs using three parameters, clinical applicability, linguistic relevance, and cultural sensitivity. Content validity index as a method for measuring content validity with a criterion set for 80% was used. This method measures the percentage of responses 6 (agree) and 7 (strongly agree) in the questionnaire. The criterion was set to consider a subtest to show acceptable content validity if 80% of the SMEs rated it 7 “*strongly agree*” or 6 “*agree*.”

The following research questions and hypotheses were addressed in this study:

1. Do the pictures used in the JAAT correctly depict the target words and sentences? Hypothesis: Item appraisals from electronic Questionnaire 1 will show 90% or more agreement on the illustrated items of the test.
2. Are there any outliers amongst the items of the JAAT? Will there be consistent errors on specific items by 30% or more of NBD examinees? Hypothesis: Most JAAT takers will not answer incorrectly with more than 20% consistency on specific items.
3. Are the items of the JAAT culturally and linguistically sensitive to the Jordanian Aphasia population? What will be the impressions of the SLPs regarding the linguistic and cultural sensitivity of the JAAT’s subtests? Hypothesis: SLP participants for electronic Questionnaire 2 will judge the

JAAT's content to be linguistically and culturally sensitive for Jordanian PWAs.

4. Can the JAAT's content be considered clinically useful for application with Jordanian-Arabic PWA? Hypothesis: The SLP participants for electronic Questionnaire 2 will judge the majority of the JAAT's content to be the clinically applicable for a tool targeting Jordanian Arabic PWA.

CHAPTER II

REVIEW OF LITERATURE

This literature review is divided into six sections. The first section will briefly provide the clinical definition of aphasia and describe its subgroups and epidemiology. The second section is a review of the three types of assessments typically used to assess language abilities of persons with aphasia (PWA): screening, functional, and comprehensive assessments. It provides examples of popular tests in English for each type of assessment. The next section discusses the advantages and disadvantages of translating and adapting existing English tests for other populations. The fourth section provides a review of the aphasia assessments currently available for Arabic speakers. The fifth section reviews content validation of tests with an emphasis on expert judgment of assessment tools and in the sixth final section, the development of the JAAT is described.

Aphasia

Definition

Aphasia is an acquired language disorder that affects communication skills after brain damage caused by cerebrovascular accident, a traumatic brain injury, or a brain pathology (Spreeen & Risser, 2003). Aphasia is usually described as a symbolic processing disorder that involves difficulty formulating and interpreting linguistic code (Hallowell & Chapey, 2008). The American Speech-Language-Hearing Association (ASHA, 2019a) defines aphasia as a communication disorder that results from damage to

the parts of the brain that contain language, typically in the left hemisphere of the brain. This disorder involves varying degrees of impairment in four primary areas: (a) spoken language expression; (b) naming or word retrieval; (c) repeating words and/or sentences; and (d) comprehension of spoken language. Primary symptoms of aphasia are impairment of the ability to retrieve the names of objects, reduced verbal fluency, difficulties in understanding spoken language, and difficulties in repeating words and sentences (Kempler, 2005). Other concomitant disorders may include reading and writing impairments (ASHA, 2019a).

Subgroups

Generally, there are two major categories of aphasia, fluent and non-fluent (Davis, 2007). Since the 1960s, aphasia has been categorized based on Norman Geschwind's groundbreaking work on localization of arterial sites of lesion (Catani & Mesulam, 2008). This work resulted in classifying aphasia into several types: Broca's aphasia, Wernicke's aphasia, global aphasia, transcortical aphasias, and conduction aphasia. Broca's aphasia is characterized by non-fluent spontaneous speech repetition with relatively intact verbal comprehension (Coppens, 2016). This non-fluency typically includes reduced phrase lengths, impaired intonation and articulation, reduced frequencies of word productions per minute, and agrammatic sentence production (Hillis, 2007). Wernicke's aphasia is distinguished by fluent yet meaningless spontaneous speech and repetition with relatively impaired comprehension of words, sentences, and conversation (Dronkers & Baldo, 2010). The type of spoken language in Wernicke's aphasia is typically limited to jargon while the person is unaware of his or her errors (Hillis, 2007). Global aphasia refers to an

encompassing deficit affecting all the areas impaired in both Broca's and Wernicke's aphasias. It is characterized by severely impaired comprehension of words, sentences, and conversation, with severely limited to no production of spoken language (Coppens, 2016). The severe lack of expressive output affects naming and reduces repetition to perseverated utterances (Hillis, 2007). Transcortical aphasias refer to syndromes similar to the above but with relatively normal ability to repeat words and sentences. Finally, conduction aphasia is denoted by fluent, accurate spontaneous speech that is phonemically similar to the intended words with a specific impairment in repetition of words (Hillis, 2007).

More recently, there has been a slow decline in the adoption of theories of localization and their characterizations of deficits in speech and language. This is due to emerging new evidence of contradictory indications (Willmes & Poeck, 1993). The recent theories of reading, naming, and sentence comprehension and production support that these tasks involve many distinct cognitive processes that could depend on different brain regions other than those classically identified as speech and language regions (Hillis, 2007).

Epidemiology

The overwhelming clinical condition leading to aphasia is cerebrovascular accident (CVA) (Croquelois & Bogousslavsky, 2011). It is well established that the frequency of individuals with aphasia among individuals with stroke ranges from 21% to 42% (Brust, Shafer, Richter, & Bruun, 1976; Kauhanen et al., 2000; Pedersen, Stig Jørgensen, Nakayama, Raaschou, & Olsen, 1995; Ryglewicz et al., 2000; Wade, Hewer,

David, & Enderby, 1986). Although these variations in frequency may be attributable to variations in methodology, sample sizes, CVA type, and medical settings (Flowers et al., 2016), more recent studies have produced comparable results (Croquelois & Bogousslavsky, 2011; Engelter et al., 2006; Inatomi et al., 2008; Kadojić et al., 2012; Stipancic, Borders, Brates, & Thibeault, 2019). An analysis of hospital discharge rates in eight American states found the percentage of discharged adult patients ($n=152,972$) having aphasia after stroke to be 18.4%. The rate of aphasia was highest among Whites compared to Blacks, Hispanics, and other ethnicities, with 77% of the sample aged 65 and older, and 56% of the sample female (Ellis, Hardy, Lindrooth, & Peach, 2017).

Between 1997 and 2006, the number of individuals in the United States diagnosed with aphasia was approximately 100,000 per year; most of them were 65 years and older, female, and lived in the South (Ellis, Dismuke, & Edwards, 2010). Aphasia individuals having suffered a CVA is linked with increased mortality, low rates of functional recovery, and low probability to regain vocational status compared with CVA-suffering individuals with non-aphasic symptoms (Engelter et al., 2006). Additionally, female gender, diabetes, and heart disease were independent predictors of aphasia (Kyrozis et al., 2009).

Prevalence rates of CVA are lower in Middle Eastern countries compared to developed countries (Feigin, Mensah, Norrving, Murray, & Roth, 2015; Krishnamurthi et al., 2015); while incidence of CVA in the Middle East comparable of those in the Western countries with reported rates ranging from 29.8 per 100,000 people in Saudi Arabia to 57 per 100,000 people in Bahrain (Tran, Mirzaei, Anderson, & Leeder, 2010).

Higher risks of CVA are usually associated with an increase in unfavorable lifestyle such as reduced level of physical activity and mounting consumption of fat and sugar-rich diets (Krishnamurthi et al., 2015). Although there are limited epidemiological studies conducted in the Middle East, El-Hajj, Salameh, Rachidi, & Hosseini, (2016) concluded that the younger population in the region indicates that CVA will become a burden in the coming years. To date, there are no published epidemiological studies providing information regarding the prevalence and incidence of CVA or aphasia in Jordan.

Assessments Used to Assess Aphasia

Aphasia testing is the initial and essential part of the rehabilitative process of individuals with communicative impairments post brain injury and many speech-language pathologists worldwide use aphasia tests in their daily professional lives. There are only a few aphasia tests in the majority of the world's commonly spoken languages and only a few tests have been standardized and normed and represented with supportive psychometric data indicating reliability and validity (Ivanova & Hallowell, 2013). Specialized clinical care and research related to aphasia are in early developmental stages in much of the world therefore, many aphasia tests in languages other than English are translations of well-known and widely used assessment instruments in English (Ivanova & Hallowell, 2013). Noteworthy examples of tests originally developed, standardized, and normed in non-English languages are the Standard Language Test of Aphasia (SLTA) in Japanese (SLTA Committee, 1977), the Aachen Aphasia Test (AAT) in German (Huber, Poeck, Weniger, & Willmes, 1983) and the Verb and Sentence Test (VAST) in Dutch (Bastiaanse, Maas, & Rispens, 2000).

In a general sense, test standardization is a term used to describe how psychometric data collected from a test is unified to ensure comparability of results between test takers. In a more specific sense, standardized refers to a uniform administration protocol which include clear instructions on how various tasks are to be administered from examinee to examinee and from one examiner to another (Ivanova & Hallowell, 2013). Thus, standardization is a process of structuring a test to unify its administration procedures, scoring system, and interpretation of scores across (ASHA, 2019b). This process is accomplished by a line of specific procedures leading to rendering the test ready for publication.

There are three major types of standardized tests commonly used for assessing PWA—screening tests, functional communication tests, and comprehensive aphasia tests. There are additional tests designed for specific linguistic aspects and tests specifically designed to test subgroups within PWA. Tests for specific linguistic aspects measure in detail one area of language performance such as naming or auditory comprehension. The Boston Naming Test (BNT; Goodglass & Kaplan, 2000) and The Peabody Picture Vocabulary Test-III (PPVT-III; Dunn & Dunn, 1997) are popular examples. Tests made to examine subgroups of PWAs focus mainly on people with severe impairments. Two examples of such tests are the Assessment of Communicative Effectiveness in Severe Aphasia (ACESA; Cunningham, Farrow, Davies, & Lincoln, 1995) and the Boston Assessment of Severe Aphasia (BASA; Helm-Estabrooks, Ramsberger, Morgan, & Nicholas, 1989).

Screening Tests

Aphasia screening tests are quick and simple tools that can indicate if a disorder is present (El et al., 2017). These tools are extremely useful clinically as they are quick and easily administered instruments for assessing PWA (Brady, Kelly, Godwin, Enderby, & Campbell, 2016) and they also provide the necessary prognostic information (Lazar, Speizer, Festa, Krakauer, & Marshall, 2008; Plowman, Hentz, & Ellis, 2012). While acknowledging all that, at their core screening tests are made to orient clinicians towards further in-depth investigations necessary for intervention (Spreeen & Risser, 2003). This is especially true for individuals showing inconsistent or difficult to categorize responses on screening tests (Ross & Wertz, 2004). Additionally, the preliminary acute communicative state of PWA, which happens when the client is first hospitalized and symptoms are prominent in severity, is subject to rapid changes which require repeated investigation after transfer or discharge (Dickey et al., 2010; Flowers et al., 2016). Examples of screening tests in English are the Bedside Evaluation Screening Test (BEST-2; West, Sands, & Ross-Swain, 1998) and the Frenchy Aphasia Screening Test (FAST; Enderby, Wood, & Wade, 1987).

Functional Aphasia Tests

Functional communication aphasia tests assess everyday communication skills to predict social functioning with an emphasis on quality of life (Sarno, 1997). Most functional communication assessment tests generally target basic communication skills, managing activities of daily living, and performing basic social skills (Hegde & Freed, 2011). This type of aphasia assessment does not provide a substantial and comprehensive

linguistic profile necessary for treatment and defining outcome measures; they play more of an additional role in the overall assessment process aimed at testing specific aspects of functioning (Howard, Swinburn, & Porter, 2010). The American Speech Language Hearing Association's Functional Assessment for Communicative Skills in Adults (ASHA FACS; Frattali et al., 1995) is a well-known example of a functional communication test.

Comprehensive Aphasia Tests

Comprehensive aphasia batteries are helpful in providing a standardized measure that attempts to provide a wide-ranging comprehensive profile of PWA' abilities, and to detect and evaluate their impairment (Bruce & Edmundson, 2010). This type of aphasia test seeks to obtain diverse samples of performance at different levels of task difficulties over all language aspects the test developer deems relevant (Spren & Risser, 2003). Unlike other types of standardized aphasia tests, comprehensive aphasia tests examine several aspects of language function such as auditory comprehension, verbal fluency, naming, repetition, reading, and writing (Papathanasiou, Coppens, & Potagas, 2013). These examinations are conducted via a reasonably broad sampling of items (Spren & Risser, 2003). Additionally, comprehensive testing batteries allow for estimation of severity of the disorder within a range of PWA and can be used to allocate treatment procedures and monitor recovery (David, 1990). Examples of comprehensive tests in English are the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972, 1983; BDAE-3; Goodglass, Kaplan, & Baressi, 2001), the Western

Aphasia Battery (WAB; Kertesz, 2006), and the Comprehensive Aphasia Test (CAT; Swinburn et al., 2005).

The BDAE is used extensively for clinical and research purposes and has been adapted for use in several languages (Tsapkini, Vlahou, & Potagas, 2010). The BDAE consists of 40 subtests divided into five sections: (a) conversational and expository speech; (b) auditory comprehension; (c) oral expression; (d) reading; and (e) writing (Spreen & Risser, 2003). The third version has subtests that include narrative story retelling, auditory comprehension, repetition, naming, reading, writing, and praxis (Goodglass et al., 2001). One of the issues that critics of the BDAE have raised is that it is heavily built on classical aphasia syndromes and that 70% of PWA cannot be categorized into that classical typology (Prins & Bastiaanse, 2004). All lesion-deficit classifications assumed that there is an invariant relationship between anatomy and function (Fridriksson et al., 2018; Rutten, 2017). The use of these classifications in diagnosis had been vigorously challenged (Byng, Kay, Edmundson, & Scott, 1990).

Howard et al. (2010) outlined several factors that justify the need for comprehensive aphasia tests:

1. Comprehensive aphasia tests provide the clinician with many benefits such as a summary of PWA' linguistic abilities and impairments and a baseline for designing language intervention.
2. A standardized comprehensive test can provide an accurate method to compare performances and measure progress by analyzing the strengths and

weaknesses in language and share the information with other clinicians and with the rehabilitation team.

The more available several tests are for a population, the better the opportunity for the clinician to collect more information. Having comprehensive tests along with other types of formal tests can fill most gaps in the assessment process. There are new assessment tools that can be used along with comprehensive aphasia tests to create in-depth portfolios of the communicative state of PWA. Some of those tools are designed to assess real-world spoken language in PWA such as the Verbal Activity Log (VAL; Johnson et al., 2014). Other tests assess speech production intelligibility in PWA, such as the Chapel Hill Multilingual Intelligibility Test (CHMIT; Haley, Roth, Grindstaff, & Jacks, 2011). Current applications of computer software in assessing PWA can also be utilized in aphasia testing (Husmann et al., 2012) and examples of such tools are the Aachen Speech Analysis (ASPA; Huber, Grande, & Springer, 2005), which is a computer-assisted method for the quantitative analysis of spontaneous speech. Also, the CommFit (Brandenburg, Worrall, Copland, Power, & Rodriguez, 2016), a smartphone application that measures talk time for PWA.

Translating and Adapting English Tests

The availability of standardized aphasia tests in Arab-speaking countries of the Middle East is very limited. One way of overcoming this limitation is by translating a standardized English aphasia test into Arabic with certain modifications. Test translation is a popular term that is used to describe the process of translating an already existing test to serve a different linguistic population. Hambleton, Merenda, and Spielberger (2004)

suggested using the term “test adaptation” for a more accurate and reflective description of what should happen in practice when preparing a test designed in a certain language and culture to be used with a linguistically different population. Adapting a test can be a less expensive option than constructing a new one and also can be useful for cross-cultural research purposes (Hambleton & Patsula, 1999), but the effort involved in test equivalence between languages has proven to be laborious and time consuming (Bridges, 2004).

van de Vijver and Hambleton (1996) described three types of biases that threaten the adequacy of translating tests: (a) construct bias that is related to non-equivalence of constructs cross-culturally, (b) methods resulting from test administration issues that may be perceived differently in distinct cultures, and (c) item bias that is directly linked to incorrect word choice during translation. Bridges (2004) states that test translation as a process is no longer mere translation; it is a process that requires careful planning and vigorous work to the extent that a test developer may consider designing his/her own test for a targeted group before committing to such a feat. This is especially important when facing two vastly distinct populations in terms of language and culture; thus, developing a new test may be the best course of action to take (Hambleton & Patsula, 1999).

Arabic Aphasia Tests

There are currently five Arabic aphasia tests available for Arabic-speaking Middle Eastern countries. One aphasia screening test available in Lebanon (A-DAB-1; Al-Thalaya et al., 2018), one functional communication test translated in Saudi Arabia (ALFA; Al Yaari & Almaflehi, 2013), one specific linguistic aspects test adapted in

Saudi Arabia (OANB; Alyahya & Druks, 2016). Two comprehensive aphasia tests are available, the first was developed in Egypt (CAT; Abou El-Ella et al., 2013; KAAT; Hassanein et al., 2002), and the other test was developed in Qatar (QACAT; Khwaileh et al., 2016). The bilingual aphasia test developed for Jordanian PWA (Paradis & El Halees, 1989) was never referenced on normal speakers or PWA in Jordan (Ivanova & Hallowell, 2013). Additionally, the test is currently not published for clinicians to purchase and use; therefore, this test is not currently available.

In the following section the six available aphasia tests in Arabic will be reviewed with the exception of the KAAT (Hassanein et al., 2002) due to the lack of available literature needed to determine its theoretical basis, type, and availability.

The Arabic Diagnostic Aphasia Battery (A-DAB)

The A-DAB is a bedside tool based on the theoretical framework of the revised version of the revised version of the WAB (WAB-R; Kertez, 2006). The WAB-R is a comprehensive testing battery that assesses seven areas of communicative functioning—spontaneous speech, auditory comprehension, naming, repetition, reading and writing, praxis, and construction and takes 45-60 minutes to administer (Kertesz, 2006). Added to this was a bedside assessment that includes extensive use of cards and pictures and takes significantly less time to administer than the whole test. To develop the A-DAB, a translation of the bedside version of the WAB was conducted and the structure of the original test was adapted and modified based on Lebanese Arabic syntax and morphosyntax structure (Al-Thalaya et al., 2018). The ADAB contains six subtests—a descriptive speech subtest consisting of three brief questions, a picture description subtest

that contains a short story illustrated by pictures, an auditory comprehension subtest that contain 10 yes/no questions, a complex auditory comprehension subtest containing four commands varying in difficulty, a repetition subtest with five Arabic words and sentences, and a naming subtest with 20 pictures of familiar objects. The developers of the test estimate the time of a complete administration of the test to be 15-20 minutes. Norming and validation studies were conducted using 60 non-brain-damaged (NBD) individuals and 30 PWA.

The Assessment of Language-related Functional Activities (ALFA)

The Arabic version of the ALFA (Al Yaari & Almaflehi, 2013) is a translation of the original English-based test (Baines et al., 1999). The original ALFA is a functional assessment therefore, not a tool built for providing a diagnosis for classifying aphasia (Spreen & Risser, 2003). The test takes 30-90 minutes to complete the 10-subtest that assess language-related functional skills—telling time, counting money, addressing an envelope, solving daily numerical problems, writing a check and balancing a checkbook, understanding medicine labels, using a calendar, reading instructions, using a telephone, and writing a phone message (Baines et al., 1999). There are no indications that Al Yaari and Almaflehi (2013) have performed any adaptation techniques to the test for the Saudi Arabian population. The authors mentioned that their translation was edited by experts of the Arabic language but specified no credentials or evidence of language adaptation for the translation. The translated version of the ALFA was then used to conduct the reliability and validity studies using 100 PWAs (Al Yaari & Almaflehi, 2013).

The Object and Action Naming Battery (OANB)

The OANB is a specific linguistic aspect test that was developed to assess confrontation naming of objects and actions in research and clinical populations (Druks & Masterson, 2000). The OANB consists of 162 black and white object pictures and 100 action pictures and was originally created to investigate the differences between naming nouns and verbs (Masterson & Druks, 1998). Adapting the OANB for use by Saudi Arabian Arabic speakers involved collecting name agreement data and values for the psycholinguistic features such age of acquisition, spoken-word frequency, imageability, and visual complexity from 30 Saudi Arabic speakers (Alyahya & Druks, 2016).

The same 262 pictures were used with the Saudi-Arabic speaking sample and pictures with less than 93% agreement were excluded from the adapted version of the OANB which eventually had 50 object pictures that achieved 100% name agreement and 50 action pictures that achieved a minimum of 93% name agreement (Alyahya & Druks, 2016). 140 Saudi-Arabic speakers were interviewed to obtain ratings for spoken-word frequency, imageability, age of acquisition of nouns and verbs, and visual complexity of the pictures that were finally included in the adapted version of the OANB.

The Bilingual Aphasia Test (BAT)

The initial purpose of constructing the BAT (Paradis & Abidi, 1987) was to compare two languages of a bilingual individual with aphasia to determine if one language might be better preserved than the other (Paradis, 2011). The test seeks cultural and linguistic equivalency of test items instead of mere translations of each by openly providing a description of the rationale used in the construction of the items of the test

(Paradis & Libben, 1987). The BAT consists of three parts: (a) a history of bilingualism questionnaire, (b) a language-specific test, and (c) a test for each specific language pair. The BAT has subtests that assess auditory comprehension of word and sentence levels, repetition of words and sentences, naming, and fluency by assessing sentence construction (Ivanova & Hallowell, 2009); supplementary testing of reading and writing is also available. The Jordanian-Arabic version of the BAT followed the development criteria set by Paradis and Libben (1987), but the test was never normed on Arabic speakers with or without aphasia (Ivanova & Hallowell, 2013). Additionally, there are no copies of the test available for purchase in Jordan.

The Egyptian-Arabic Comprehensive Aphasia Test (E-CAT)

The CAT is a comprehensive aphasia test that provides a profile of performance across several modalities of language production and comprehension. The language battery consists of 34 subtests and it also includes a cognitive screen and a disability questionnaire (Swinburn et al., 2005). The language battery is designed to assess language comprehension, repetition, spoken language production, reading and writing. Abou El-Ella et al. (2013) modified the CAT (Swinburn et al., 2005) for the Egyptian-Arabic speaking population by translating the test into Arabic and modifying some of its subtests to be culturally suitable to the Egyptian environment. Modification of the CAT was extensive; an unspecified number of pictures were replaced by pictures more familiar to the Egyptian environment within all subtests (e.g., word and sentence identification, naming, and picture description). Several words, sentences, and paragraphs were changed according to the frequency of occurrence of the words in Arabic and the grammatical

structure of Arabic (Abou El-Ella et al., 2013). No clarification of the methodology used to specify the frequency of occurrence of the Arabic words and sentences was offered. The standardization of the E-CAT was based on the results of test administration to 50 NBD Egyptian Arabic-speaking individuals and the set of 100 test results from PWAs (Abou El-Ella et al., 2013). The test modification team used the same standardization data to investigate the reliability and validity of the E-CAT.

The Qatari-Arabic Comprehensive Aphasia Test (Q-CAT)

The Q-CAT is a project that is in line with the CAT's (Swinburn et al., 2005) design as described by its developers (Khwaileh et al., 2016). The test development effort was carried out in four stages: (a) a review of Qatari-Arabic linguistics and psycholinguistics (e.g., name agreement, imageability, age of acquisition, image agreement, and familiarity) was conducted to identify elements to be used in the test; (b) a normative database of lexical stimuli of pictures and words that includes 530 nouns, 250 verbs, and 150 adjectives was collected from 160 Qatari speakers; (c) a set of aphasia subtests was created using normative database, including a cognitive screen, a language battery, and a disability questionnaire similar to the structure of the CAT; (d) data from 50 healthy adult Qatari-Arabic speakers was used to generate control data for the battery subtests to create normative scoring; and (e) test was administered on Qatari-Arabic PWA (Khwaileh et al., 2016). The test is to be published for clinical and research use in August of 2019.

Content Validity

Tests are popular tools that are used for clinical, educational, research, and policymaking purposes. The process of developing a test, using the test for certain purposes, and evaluating a test cannot occur without the utmost consideration of validity. Putting it simply, validity is the degree with which a test actually assesses what it was designed to assess (Garrett, 1937). This simplistic view is largely refuted today—although it is still seen in many textbooks—and is considered an incomplete definition (Sireci, 2009). To provide more organization to the state of test design procedures, the American Psychological Association (APA) in collaboration with the American Educational Research Association (AERA) and the National Council on Measurement in Education (NCME), published the Technical Recommendations for Psychological Tests and Diagnostic Techniques in 1954 (Anastasi, 1986). Validity was subdivided based on approaches of investigating validity that eventually came to be known as “types” of validity. Those kinds or types of validity were documented in the statements made by joint committees and distinguished several types of validity including content, predictive, status, and congruent (Sireci & Parker, 2006). The types of validity were eventually reduced to three—content, criterion, and construct (APA, AERA, & NCME, 1966).

It is important to differentiate between validity and validation. Validation is the process that examines the degree of validity; therefore, validity can be described as the level of adequacy of inferences derived from a test’s scores, and validation is the action of assessing the degree of adequacy (APA, AERA, & NCME, 2014). Content validity is built on professional judgment of the relevance of a test’s content to the content of a

behavior domain and how well the items/tasks of the test represents the content domain in question (Messick, 1989). In certain ways, content validity gives us an almost direct link between the procedures presented in the criterion measure and the derived interpretations or uses of test scores (Cureton, 1951; Ebel, 1961). The judgement process, the most prominent goal of content validation, involves asking a number of experts to evaluate the validity of items and/or the entire assessment to retain the best aspects that adequately measure a desired content domain within a population (Tojib & Sugianto, 2006). Used well, there is evidence that review by SME can be very effective and relatively inexpensive (DeMaio & Landreth, 2004; Irwing & Hughes, 2018; Presser & Blair, 1994). Content validity research has an important role in the development and testing processes for any instrument and should be a priority during the development process because it is a prerequisite for evaluating other types of validity (Slocumb & Cole, 1991). The recruitment of experts to review and provide critique for content of an instrument should be based on predefined criteria that consider the qualifications, experience, and clinical experience of SMEs (Grant & Davis, 1997).

Although content experts' reviews of assessment tools are used in professional, educational, and psychological testing (Mishra, Catchpole, & McCulloch, 2009; Penny, Waschbusch, Klein, Corkum, & Eskes, 2009; Schmidt et al., 2009; Seo, MacEntee, & Brondani, 2015), such evidence of content validity is seldom obtained and/or presented, although it would be highly desirable (Carretero Dios & Pérez, 2007; Delgado-Rico, Carretero Dios, & Ruch, 2012). There is no evidence referring to the use of this method of content validation in major English language comprehensive aphasia tests. There are

no studies that examined the content validity of the BDAE-3 (Goodglass et al., 2001), and in the case of the WAB (Kertesz, 2006), the test designers used the fact that the WAB is a modification of the BDAE to establish grounds for content validity (Davis & Finch, 2010). The test designers discussed how the similarity to the BDAE proves that the WAB meets subjective criteria for content validity by the fact that BDAE is highly correlated with other standardized tests (Acheston, 2010). The manual of the CAT (Swinburn et al., 2005) included data on reliability and certain types of validity but contained no information on content validity. Within Arabic tests for PWA, the designers of the DAB-1 (Al-Thalaya et al., 2018) used the feedback of eight Lebanese SMEs on each stimulus of the test.

The Jordanian Aphasia Assessment Test (JAAT)

The number of aphasia assessments for Arabic speakers in the Middle East is inadequate as there are, for example, no comprehensive aphasia assessments for Jordanian Arabic speakers and no published literature to suggest the availability of such tests in Arabic-speaking countries other than Egypt and Qatar. Tunisia, Libya, Syria, Iraq, and Algeria are also without suitable aphasia tests. If a potential test developer is facing this problem, what type of aphasia test should take priority?

As discussed earlier, comprehensive tests offer more diagnostic information by assessing a broad set of certain linguistic aspects that are commonly affected by aphasia. Other types of aphasia tests cannot emulate the amount of construct covered by comprehensive aphasia tests. These attributes make it very hard to dispute the crucial need to use a comprehensive test in a clinical setting (Marshall & Wright, 2007).

The BDAE is currently one of the most commonly used tests (Wilson, Eriksson, Schneck, & Lucanie, 2018) and has proven influential in creating another comprehensive test, the WAB (Al-Thalaya et al., 2018). The BDAE is a large test that takes 90-120 minutes to administer the standard version, while the extended version takes up to 2 and one-half hours (Sbordone, Saul, & Purisch, 2007); this makes the BDAE the most extensive battery available for assessing PWAs. Due to the lack of comprehensive language tests for PWA, a team of five SMEs lead by the author of this study developed an aphasia test for the Saudi and Jordanian PWA in 2012. At that time the aphasia examination was named the Arabic Aphasia Test (AAT) and it was not published or used clinically. The AAT was later modified and further developed by the author between the years 2016 and 2018 to focus exclusively on Jordanian-Arabic speakers. The test was named the Jordanian Arabic Aphasia Test (JAAT) and it is markedly different from the AAT. The JAAT is structurally influenced by the BDAE but like the WAB before it, the JAAT does not borrow any linguistic or visual items from the BDAE and does not translate or adapt any of the BDAE's stimuli. In the following section, a description of the subtests that were influenced by the BDAE is provided. This description of influence follows the creators of the WAB's (Kertesz, 2006) argument which used the fact that the WAB is a modification of the BDAE to establish grounds for content validity (Davis & Finch, 2010), but unlike the WAB, the development of the JAAT did not stop there.

Adaptations of tests of aphasia in Arabic depend heavily on materials translated from English (e.g., El-Ella et al., 2013) that are specifically developed for English speakers. These adaptations are driven by the structure of English and neglect linguistic

properties of the Arabic language (Khawaileh et al., 2016). The structure of the JAAT is similar to the BDAE, as it contains four parts—verbal fluency, auditory comprehension, naming, and repetition but it attempts to account for linguistic influences.

As discussed above, the JAAT had its original basis from the AAT, an early version that was developed by five subject matter experts (SMEs) in 2012. The preliminary aim was to design a test that could be used for Jordanian and Saudi PWA. The Saudi and Jordanian populations have tens of thousands of common words and the two population have relatively no difficulties in verbal communication (Saidat, 2018). The group consisted of certified Arabic-speaking speech-language pathologists (SLPs)—one SLP had a doctoral degree, one a master’s degree, and three with bachelor’s degrees. These SLPs each had at least 5 years of clinical experience assessing and treating PWA. Construction of the AAT involved selecting target words and sentences based on a rating system of familiarity, morphologic complexity, and imageability. Target words were categorized based on SMEs’ agreement on these categories and later listed for inclusion in the test. The next stage involved creating culturally appropriate depictions of words, sentences, and paragraphs for the Saudi and Jordanian cultures. Several artists were recruited for the visual work and one artist was eventually hired based on drawing the most salient visual representations of the selected linguistic stimuli. The JAAT is a heavily modified version of the AAT and has seen many changes in the past four years by the author of this study to even more accommodate the Jordanian-Arabic population. The JAAT was designed as a different test with more than 35 items changed using the efforts of a new artist from Jordan to redraw all pictures for increased saliency. For example, 27

words and pictures used in the word comprehension subtest were taken from the Khwaileh, Body, and Herbert Levantine Arabic normative database (2014). The JAAT is still in its phase and no version of it has been published.

Elements of the JAAT

The JAAT (Appendix A) is divided into four parts based on the four major linguistic abilities that highlight aphasia—verbal fluency, auditory comprehension, naming, and repetition of words and sentences. The JAAT has 12 subtests containing a total of 112 items. All 12 subtests have specific instructions and scoring guidelines on the answer sheet. Some subtests require practice items to decrease the confounding factor of unfamiliarity. Practice items are introduced in the following subtests: single-word picture identification, sentence-picture identification, sentence and paragraph comprehension, responsive naming, and categorical naming. The construction and content of the JAAT is described in the following sections.

Verbal fluency. There are three subtests in this section: response to open-ended questions and two picture descriptions tasks. *Open-ended questions* have 10 items with 10 points as the maximum score. The subtest *Picture description* and *Picture description of sequences* each has a maximum score of six points, making the total score of this part of the JAAT 22 points.

Open-ended questions. This subtest consists of 10 open-ended questions about daily living (e.g., How are you today? What is your full name? What happened to you?). Four of those questions were chosen because of their cultural equivalency and six new questions were designed by the team (e.g., Who brought you here? What do/did you do

for a living?). Questions were simple clauses with familiar words that focused on daily living activities. One point is given for each appropriate social response making the total score for this subtest 10. This subtest can be found in Appendix A, Part 1.

Picture description. Almost all major formal comprehensive aphasia tests use picture description tasks for eliciting verbal responses (Goodglass et al., 2001; Kertesz, 2006; Swinburn et al., 2005). The pictures used contain several actions within a picture. These pictures usually reflect the cultural settings of the target populations. The JAAT team discussed certain visual scenarios and opted to use the familiar social event in Jordan of a guest room with visitors. This subtest consists of one picture with routine actions depicting real-life events in *a guest room* (see Appendix A, Part 1). For scoring purposes, the picture was divided into six major event parameters: (a) girl holding teddy bear/toy; (b) cat's tail being stepped on by girl; (c) man holding drinks; (d) drinks are spilling onto guest; (e) guest wearing traditional attire is reading a book; and (f) boy is playing with a lighter. One point is given for successful verbal production of each event parameter making the maximum score for this subtest six points. Participants are asked to describe the picture in detail. For participants who missed an event, the examiner points to the event and asks for a description. The final picture was revised based on input from faculty and doctoral students in Communication Sciences and Disorders (CSD) in the United States.

Picture description of sequences. This subtest is similar to the *narrative discourse* subtest in the extended version BDAE-3. The BDAE-3 subtest consists of several illustrated cartoons that tell a short story. This subtest is available in the extended

version of the BDAE-3 because it takes a long time to administer. In the JAAT, this subtest consists of two images of *a woman in the kitchen* (Appendix A, Part 1). The two pictures contain two major events and actions are described on the answer sheet. A similar scoring procedure explained in the previous picture description subtest is used for calculating the final score. The pictures are divided into six major event parameters: (a) woman in kitchen; (b) woman putting chicken in oven; (c) woman speaking on phone; (d) woman is distracted; (e) smoke is coming out of oven/food is burning; and (f) participant is to name at least two items appearing on the kitchen table. One point is given for successful verbal production of each event parameter, making the maximum score of this subtest six points. Participants are asked to describe the picture in detail. For examinees who miss an event, the examiner points to the event and asks for a description.

Auditory comprehension. The auditory comprehension part has four subtests: *Comprehension of commands*, *Comprehension of single words*, *Comprehension of sentences*, and *Sentence and paragraph comprehension*. The maximum score for the four subtests is 47: Comprehension of commands: 9 points; Single-word comprehension: 20 points; Sentence-picture identification: 6 points; and Sentence and paragraphs comprehension: 12 points.

Comprehension of simple commands. This subtest is called *Commands* in the BDAE-3 and has only five commands. The team decided to include nine items because following instructions is important in rehabilitative settings. All commands contain everyday actions and concrete words. This subtest assesses the comprehension of speech instructions varying in complexity. There is a total of 9 items; three one-step commands

(e.g., *close your eyes*), three two-step commands (*hold the pen and write your name*), and three three-step commands (e.g., *hold the pen, draw a line, and put the clock on the paper*) (Appendix A, Part 2). Five familiar objects (pen, watch, small book, white sheet of paper, and a tissue/kleenex) are used in multistep instructions. One point is given for following each consecutive set of commands. Any attempt is deemed erroneous and had a score of zero if a single command was not followed within an item. Nine items gradually increasing in number of commands resulted in a total possible score of 9 on this subtest.

Single word comprehension. The BDAE-3 has 20 items depicting words that are distributed equally between three categories (tools, foods, animals, and actions). To reduce time needed to perform the test, it was decided to have 20 words that reflected the following categories and items: tools/personal items (hammer, nails, roller, wallet), animals (giraffe, frog), foods (apple, pineapple), and actions (eat, count, drive). Foil pictures had visual, phonemic, and/or syntactic similarities with the target word. For example, the word نار or *fire*, is spelled *Naar* in Arabic, a visual distractor will be a fountain, a phonemic distractor is فار or *Faar*, the semantic will be حطب or *Firewood*. A set of words was created by each team member. Each word was rated for cultural appropriateness, familiarity, and how easy it was to visually depict and create foils that were visually, semantically, and/or phonologically similar. Ten items were created from the Levantine normative database (Khwaileh et al., 2014).

In this subtest, each item had four pictures, one picture represents the target word and the other three pictures are distractors or foils. The three foils are visually or phonologically similar to the target stimulus. One point is given for each correct response

produced in less than five seconds and half a point for correct response produced after five seconds. Appendix A, Part 4 contains the items in this subtest.

Simple and compound sentence comprehension. This subtest is similar to the *Embedded Sentences* subtest on the BDAE-3. While the BDAE-3 uses ten pictures and simple sentences, the subtest in the JAAT contains six simple sentences with four response choices. The three foils are visually similar to the correct answer. For example, for the sentence *The cook is chasing the physician*, the three foils are a physician chasing a cook, a cook painting a wall while the physician is watching, and a physician painting a wall while the cook is watching. The following guidelines were followed to implement this rationale: (a) the word used in the sentences must be culturally appropriate; (b) the words used in the sentences must be lexically simple because the priority is assessing sentence comprehension (Swinburn et al., 2005); (c) the target words used in the sentences must be words that can be visually represented; and (d) sentences must describe various actions and social or vocational contexts.

In this six items subtest, the examiner reads sentences representing actions and events as depicted in one of the four pictures (Appendix A, Part 2): *I will now read a sentence and show you four pictures. Point to the picture that represents the sentence I read.* One point is given for correct responses. A maximum score of 6 points is given to each examinee who responded correctly on all items.

Multiple sentence and paragraph comprehension. The BDAE-3 contains four short stories with varying length and complexity with 16 questions. The JAAT has three stories with 12 questions in the interest of saving time. As with previous subtests in the

JAAT, the sentences and paragraphs were designed to be representative of Jordanian culture. The first item contains two compound sentences followed by two yes/no questions. The second item contains four sentences followed by four yes/no questions. The third item contains a paragraph with six sentences that are followed by six yes/no questions. One point is given for each correct answer. A total score of 12 is given for each participant who responds correctly to all items.

Naming. The naming assessment part of the JAAT was designed by creating a pool of target words representing objects and actions. There are four naming subtests: Object naming, picture naming, responsive naming, and categorical naming. The naming part has a maximum score of 25: Object naming, 5 points; Picture naming, 15 points; and Responsive naming, 5 points. Categorical naming had no predetermined score because there are no cutoff scores to what the maximum production can be therefore, for purposes of this study the number of words produced were reported without adding it to the final scores of the JAAT.

Object naming. The BDAE does not use objects for the naming subtest or in any other part of the test. The five objects used in the JAAT's *comprehension of instructions* subtest are pen, watch, book, white sheet of paper, and a tissue/Kleenex and one point is given for each correct response.

Picture naming. The team created 15 word/picture items from the pool of potential words. The BDAE-3 uses more than 40 pictures for the naming section including number, letter, and color naming. For time management reasons, the team

reduced the number to 15 pictures. The pictures were chosen based on (a) familiarity of word to the target population, (b) word length, and (c) lexical complexity.

Appendix A, Part 4 has examples of pictures used for this subtest. Pictures are presented one at a time. Prompts are, “What is this?” for objects and “What is he/she doing?” for actions. One point is given for each correct response.

Responsive naming. This subtest uses questions that logically end with one specific target word. As in previous subtests, the words reflect concepts tied to social or vocational settings and the questions used to elicit words are syntactically simple. The five questions in this subtest require specific one-word answers; for example, “*Where do we store milk so it can't be spoiled?*” One point is given for each correct response produced in less than 5 seconds, and half a point for correct response produced after 5 seconds.

Categorical naming. This subtest is not found in the BDAE-3 but it is found in the CAT. The involvement of semantic working memory can be useful in providing more information related to adults with primary progressive aphasia (Marczinski & Kertesz, 2006). It was added to assess the ability of PWA to retrieve words using a semantic cue and the categories tested are fruits and animals. In this subtest, the examinee is given one minute to name all the animals he/she can recall and verbally produce.

Repetition. The extended version of the BDAE-3 uses ten single words, four nonwords, and ten simple sentences with varied numbers of words within each sentence. The short version of the BDAE-3 uses five single words and two sentences. To reduce the length of the JAAT, it was decided to have ten single words, three nonwords, and five

sentences—18 items instead of 24 on the extended BDAE-3. Like the BDAE-3, the single words varied in syllabic structure (two one-syllable words, four two-syllable words, and four multisyllabic words) and sentences ranged from two to 11 words.

This fourth and final part of the JAAT has one subtest that contains 18 items: ten monosyllabic, bi-syllabic, and multisyllabic regular words, three nonwords, and five sentences. Responses were recorded and transcribed and repetition must be intelligible to be considered correct. One point is given for a correct response and no points for responses that missed one word in items 1 through 17. Item 18 was considered incorrect if two words or more were missed by the participants. All items are shown in Appendix A (Part 3).

Statement of Purpose

Few comprehensive aphasia tests exist in most the world's commonly spoken languages other than English (Fyndanis et al., 2017). This has led to the adaptation of English tests to other languages (Ivanova & Hallowell, 2013). Many clinicians believe that translating a test requires less effort in comparison to designing a new one (van de Vijver 1996), but test adaptation often does not address linguistic and cultural differences that influence performance (Geisinger, 1994; Hambleton, 1996; Hambleton & Patsula, 1998).

Although several recent efforts in the Middle East have resulted in the development of a number of assessments (Abou El-Ella et al., 2013; Al-Thalaya et al., 2018; Al Yaari & Almaflehi, 2013; Alyahya & Druks, 2016; Khwaileh et al., 2016), Jordan currently does not have a comprehensive aphasia test.

The goal of this current study was to assess the content validity of the newly developed JAAT as a primary attempt to address the need for a comprehensive aphasia test in Arabic to be used in Jordan. This validation study was carried out through three experiments:

1. Experiment One (Questionnaire 1: Verification of visual stimuli by NBD Jordanian Arabic speakers): The first step in the validation process was designed to verify that the visual stimuli used in the test was based on salient representation of the words and sentences they target. Experiment One examined whether a picture adequately represents the target word it was designed to depict. This was achieved via an electronic questionnaire (Questionnaire 1) that presented all the pictures used in the JAAT to 164 NBD Jordanian Arabic speakers (see Appendix B). The participants in this questionnaire were asked to choose the correct word that each picture represented from four single-word options. The research question for Experiment One was: *Do the pictures used in the JAAT correctly represent the targeted words and sentences? The hypothesis for experiment one was: Item appraisals from electronic Questionnaire 1 will show more than 90% agreement on most of the illustrated items of the test.*
2. Experiment Two (identifying inadequate items in the JAAT-2 by administering the test on senior NBD Jordanian Arabic speakers): The second experiment was created to pinpoint any outliers in the pool of items by administering the JAAT-2 to 20 senior NBD Jordanians to identify any items

that show 30% or more error that would be replaced or modified accordingly.

The research question for Experiment Two was: *Are there any outliers in the JAAT's items? Will there be consistent errors on specific items by more than 20% of normally speaking examinees? The hypothesis for Experiment Two was: The majority of JAAT takers will not answer incorrectly with more than 20% consistency on specific items.*

3. Experiment Three (Questionnaire 2: Validating of test content based on SME rating of the JAAT-3): Experiment Three gave the JAAT-3 to 21 certified Jordanian SLPs and asked them to appraise the test's subtests, parts, and the whole test as a unit of assessment. The SLPs were asked to participate in Questionnaire 2 the purpose of which was assessing the clinical value, linguistic relevance, and cultural sensitivity of the JAAT. The research questions addressed in Experiment Three were:
 - a. *Are the items of the JAAT culturally and linguistically sensitive to the Jordanian population with aphasia? What will be the impression of the SLPs regarding the linguistic and cultural sensitivity of the JAAT's subtests? The hypothesis for this part was: SLP participants in electronic Questionnaire 2 will judge the JAAT's content to be linguistically and culturally sensitive for Jordanian PWA.*
 - b. *Can the JAAT's content be considered clinically useful for application with Jordanian-Arabic PWA? The hypothesis for this part was: The SLP participants in electronic Questionnaire 2 will judge the majority of the*

JAAT's content to be clinically applicable for a tool targeting Jordanian Arabic speaking PWA.

CHAPTER III

METHOD

This study consisted of three experiments that were conducted in the following order:

1. Experiment One (Questionnaire 1: Verification of visual stimuli by NBD Jordanian Arabic speakers);
2. Experiment Two (identifying inadequate items in the JAAT by administering the test on senior NBD Jordanian Arabic speakers); and
3. Experiment Three (Questionnaire 2: Validation of test content based on SMEs ratings of the JAAT).

After Experiment One and Two, the JAAT was modified before conducting Experiment Three. These versions of the JAAT are denoted in this paper as JAAT-2 (the version of the test made after Experiment One modifications); and JAAT-3 (the version of the test made after Experiment Two).

Institutional Approval

Each experiment had its separate individual segment approval by the University of North Carolina at Greensboro (UNCG) Office of Research Integrity. Experiment Two and Experiment Three utilized the participation of subjects from Jordan in their respective online questionnaires. The subjects in Experiments One and Three did not

require approval from any international organization to participate in the online questionnaires.

Experiment Two (identifying inadequate items in the JAAT by administering the test on senior NBD Jordanian Arabic speakers) required the recruitment of subjects speaking Jordanian Arabic living in North Carolina. For subjects who were literate, the author of the study explained the purpose of the study and presented a consent form to them. They were asked to read consent form and sign it if they understood and approved the terms. For subjects participating in same experiment who were illiterate, the author was required to explain the purpose of the study and the consent procedure verbally to them and their families and ask their families to read the assent form to the subject and explain it individually to them before signing it.

Study Design

To address the research questions, the study used two questionnaires and one test administration on a group of non-brain-damaged (NBD) Jordanian individuals. The first questionnaire is used in Experiment One (Questionnaire 1: Verification of visual stimuli by NBD Jordanian Arabic speakers) and the second questionnaire is used in Experiment Three (Questionnaire 2: Validation of test content based on SME rating of the JAAT-3). Content validity was assessed through Experiment One and Experiment Three using the two questionnaires—one for each experiment. Experiment Two (identifying inadequate items in the JAAT-2 by administering the test on senior NBD Jordanian Arabic speakers) was designed to identify items with high incorrect responses for the purpose of modifying

or replacing test items. This was achieved by administrating the JAAT on NBD Jordanian Arabic speakers.

Experiment One (Questionnaire 1: Verification of visual stimuli by NBD Jordanian Arabic speakers) used a Qualtrics online questionnaire (Appendix B) to assess visual stimuli used in the preliminary version of the JAAT, or JAAT-1. Subjects were 164 Jordanian Arabic speakers varying in age and educational background.

After verification was achieved, JAAT-2 was made and used in *Experiment Two* (identifying inadequate items in the JAAT by administering the test on 20 senior NBD Jordanian Arabic speakers) to identify items in JAAT-2 (Appendix A) that proved to be difficult for 30% or more of subjects in the experiment.

JAAT-3 is the version of the JAAT that was created after modifications were made based on the results of Experiment Two. This version of the JAAT was used in *Experiment Three* (Questionnaire 2: Validation of test content based on SME rating of the JAAT-3) to assess the test's content based on three parameters—clinical applicability, linguistic relevance, and cultural sensitivity. This was achieved using an online Qualtrics questionnaire. The SMEs were provided with soft copies of the JAAT-3 and were advised to thoroughly review it before responding to Questionnaire 2. SMEs were not asked to use the rest on their clients.

Experiment One (Questionnaire 1: Verification of Visual Stimuli of the JAAT-1 by NBD Jordanian Arabic Speakers)

Participants

Participants were recruited by contacting leaders of public and private social organizations such as the Grand Mariam Mosque in Amman and the National Democratic

Ba'ath Party of Jordan (NDBP). The vice president of the Ba'ath party agreed to distribute flyers in the party's headquarters and the Imam of the Mariam Grand Mosque was asked via email to distribute a flyer in the Mosque. The flyer contained a brief description of the study, a link to the survey to potential participants in Jordan, and contact information of the author of this study. Social media such as Facebook and Twitter were also used extensively. Participants were directed to a link for Questionnaire 1 using Qualtrics. Consent was asked from each participant electronically online. Participants were emailed or shown via social sites a letter explaining the purpose of the study and how long it would take them to complete the questionnaire.

The initial number of participants recruited for Questionnaire 1 was 179 adults. Eight subjects were excluded after answering that they did not hold Jordanian citizenship. Participants were excluded from the study if they reported having any communication impairments. Seven subjects were excluded because they stated that they had a history of brain injury in the questionnaire. Illiterate participants were helped by individuals who aided in reading all the related written material of the questionnaire for them. The Qualtrics online link of Questionnaire 1 was available for participants to complete electronically using computing devices such as laptops, tablets, or mobile phones.

Thus, Experiment One Question 1 participants included 164 Jordanian Arabic-speaking individuals; $n=100$ were females (61%) and $n=64$ were males (39%). From the 164 participants recruited, $n=70$ were between the ages of 18 to 35 (42%), $n=68$ were between the ages of 36 to 60 (41%), and $n=26$ were aged above 60 (16%). Educational backgrounds of participants were diverse; $n=4$ were illiterate (2%), $n=12$ had 6 years of

education (7%), $n=35$ had 7 to 12 years of education (21%), $n=82$ had 16 to 18 years of education (50%), and $n=31$ had more than 18 years of education (19%).

Measures

The online Experiment One Questionnaire 1 consisted of 137 questions. The first six questions collected demographic information (e.g., sex, nationality, age, and level of education). The remainder of the questionnaire consisted of 131 questions for 121 pictured items from the JAAT (a sample of Questionnaire 1 can be found in Appendix B). This represented all illustrated items used in the test. Experiment One Questionnaire 1 pictures represented nouns, verbs, and sentences, with one question for each picture in the questionnaire. The question asked the participants to select the word that best represented the target picture from a list of four words. Completion of the questionnaire was anticipated to take about 20-25 minutes.

Procedures

Once the link is accessed, participants are shown information about anonymity and confidentiality as well as consent information. The participants were asked to complete the questionnaire containing all the pictures used in the JAAT. The questionnaire was completed online by all participants.

Data Analysis for Experiment One Questionnaire

The purpose of Experiment One was to identify outliers in the depicted images that demonstrated less than 90% agreement. Percent correct was calculated for each test image item via Qualtrics percentile ratios of correct responses. To address the first research question of the study and to be included in Experiment Two, 90% or more of the

participants must have selected the correct response for each image/word. Items scoring less than 90% correct responses were either eliminated from the JAAT-1 version or modified to increase their clarity to be used as JAAT-2 in Experiment Two.

Experiment Two (Identifying Inadequate Items in the JAAT-2 by Administering the Test on Senior NBD Jordanian Arabic Speakers)

Participants

Twenty adults aged 60 years and older participated in Experiment Two. Subjects were recruited through distribution of flyers in social and religious organizations seeking Jordanian Arabic-speaking individuals within the Jordanian Arabic speaking community in the Triad area of North Carolina (Greensboro, High Point, and Winston Salem). This method was selected as a more time-and cost-effective option compared to traveling overseas to Jordan, as the test was administered by the author of the study. The Islamic Center of the Triad (ICT) in Greensboro, North Carolina was contacted and flyers were distributed with information about the experiment and contact information for the study's author was provided. Jamesford Meadow Clubhouse used by the Arab community in High Point and Greensboro was also contacted and flyers were distributed. Social media sites were also used for recruitment. Once a participant was contacted, a cover letter was orally provided to explain the purpose of the study. A consent form was signed by participants who were literate and assent forms were signed by family members after the study was explained and consent information were read to them by their caregivers.

Participants were included in the study if they were Jordanian Arabic speakers aged 60 or above and living in North Carolina. The following questions were asked to participants and their caregivers to help with the inclusion and exclusion process: (a)

have you had any neurologic debilitating incidents such as traumatic brain injury or stroke? (b) Have you been diagnosed with any neurological disease? (c) have you had any sudden or gradual loss of memory? (d) Have you had any sudden or gradual loss in speech or hearing? (e) are you taking any neurogenic medication that alters your level of conciseness? If yes, was this medication prescribed for treating brain related issues?

Participants were excluded from the study if they or their caregivers reported having disabilities such as intellectual disabilities, hearing impairment, visual impairment, brain injury, or behavioral disabilities. All participants had no history of speech and language disorders. Testing occurred in participants homes after consent forms were signed. The study was approved by the University of North Carolina at Greensboro Office of Research Integrity

Measures

The JAAT-2. The JAAT-1 (Appendix A) is described in the Jordanian Aphasia Assessment Test section of the literature review of this study. After the pictures from the JAAT-1 were validated and modifications of two pictures made, the test was ready for use in Experiment Two.

Procedures for Experiment Two

Once signed consent and assent forms were obtained, the JAAT-2 was administered over two consecutive sessions. The JAAT-2 was administered in the first session. The second session was very brief with the participants required to only to verbally describe or name seven pictures presented to them. These pictures were presented to provide an alternative pool of items for potential use in the JAAT-3. The

second session was conducted immediately after the first session. Average time of session one was 22 minutes and 44 seconds and average time for session two was 3 minutes and 26 seconds. Overall, average assessment session time was less than one hour in average. Testing sessions were conducted through one-on-one interviews by the examiner at the participants' homes. In the first session, the JAAT-2 was administered using all four major parts of the test with their designated subtests. Two stimuli used in the first session that scored less than 90% correct response rate were replaced in the second testing session. These two stimuli were "screw" (87% correct responses) in the Naming part and "lighter held by the boy" (60% correct responses) in the Verbal Fluency part.

Administration of the JAAT-2 and its scoring were performed manually using the test's answer sheet. The second session addressed the testing of modified subtests, namely the Picture Description subtest, Picture Sequence Description subtest, and five alternative pictures to be used in the Naming part after the elimination of pictures that had lower than 80% correct responses. The second session was conducted to create alternative stimuli for the final version of the JAAT to be used in Experiment Three.

Scoring of the JAAT-2 and Data Analysis of Experiment Two

The JAAT-2 uses a numerical scoring system. All subtests have a threshold for a perfect score except for the categorical naming subtest. The verbal fluency part had three subtests with 22 points as maximum score: The Open-ended questions subtest had 10 items with 10 as a maximum score, while each of the two picture description tasks had a maximum score of 6 points. Table 1 provides a brief description of the scoring. The

purpose of this experiment was to identify items that elicit less than 80% correct responses for either elimination or modification of the items before being used in the modified version of the JAAT-2 in Experiment Three.

Table 1

Subtest Points and Scoring Methods

Subtest	Maximum points	Scoring method
Open-ended questions	10	Each correct response is scored with one point
Picture description	6	Each correct response is scored with one point
Picture description of sequences	6	Each correct response is scored with one point
Comprehension of commands	18	Each correct response is scored with one point
Single word comprehension	20	Each correct response in less than 5 seconds is scored with one point. Correct responses after 5 seconds is scored with half a point.
Sentence-picture identification	6	Each correct response is scored with one point
Sentence and paragraph comprehension	12	Each correct response is scored with one point
Object naming	5	Each correct response is scored with one point
Picture naming	15	Each correct response in less than 5 seconds is scored with one point. Correct responses after 5 seconds is scored with half a point.
Responsive naming	5	Each correct response is scored with one point
Categorical naming	N/A	Varies by examinee. No norms available
Repetition	18	Each correct response is scored with one point. Incomplete repetitions of words and/or sentences is scored with half a point.

The purpose of this experiment was to identify outliers therefore, each item was analyzed for type of response. Percentile calculations were used to identify items that had 70% or less correct responses. Each subject's overall score and scores on subtests and

parts were manually input into SPSS. Descriptive statistics calculated were means and standard deviation of scores on the JAAT-2 for documentation purposes. No additional statistical analysis was required. The SPSS data was stored on a password protected laptop. Box online storage provided by UNCG was used to store data securely. Answer sheets of the JAAT-2 were kept on a paper file in a locked cabinet at the office desk of the author.

Experiment Three (Questionnaire 2: Validation of Test Content Based on SMEs' Ratings of the JAAT-3)

Participants

The purpose of Experiment Three was to assess content validity through expert judgments of the JAAT and its subtests and parts. This was achieved via a second Qualtrics online questionnaire. Speech-language pathologists with experience working with adult PWA in Jordan provided appraisals based on three major parameters—clinical applicability, linguistic relevance, and cultural sensitivity. Ten professional leaders in the field of speech language pathology who manage private clinics and/or holding academic university positions were contacted and provided with a description of the study. These professional leaders were asked to participate online in Questionnaire 2 (validation of test content based on SME rating of the JAAT-3) and provide names and contact information of eligible clinicians who met the inclusion criteria of the study. All participants were contacted and recruited via email or cellphone and asked to complete the online Qualtrics Questionnaire 2 (validation of test content based on SME rating of the JAAT-3). Final number of participants was 21 Jordanian-Arabic speech-language pathologists (SLPs)

certified to work in Jordan and had a minimum of 2 years of clinical experience in assessing and treating adult PWA.

Participants held undergraduate and/or graduate degrees in speech language pathology and obtained valid clinical certification to work in Jordan. Participants were excluded from the study if they had less than two years of experience working with PWA or had no local clinical certification. Consents was collected electronically and cover letters explaining the purpose of the study to participants were distributed. The initial number of participants recruited for the study was 46 which was later reduced to 31 after excluding 15 participants who had less than 2 years of experience in working with PWA. The study was approved by the University of North Carolina at Greensboro Office of Research Integrity.

Measures for Experiment Three

Questionnaire 2 (Validation of test content based on SME rating of the JAAT-3). The online questionnaire was designed via Qualtrics to follow the outline of the JAAT-3. The first six questions were demographic information questions about the participants (e.g., sex, nationality, clinical certification, and number of years working with PWAs). As the JAAT-3 was divided into four major parts, each part had several subtests. For example, the questionnaire started with the three subtests contained in the first part of the JAAT, which was Verbal Fluency. After each subtest was addressed with four items (Appendix D), the whole Verbal Fluency part was addressed with two items (Appendix E). This format is repeated until all the subtests and parts were examined. After that, eight major questions that addressed the JAAT as a whole unit were presented

for participants (Appendix F). The questionnaire had 73 items. Qualtrics gave an estimated completion time of 15-20 minutes.

Procedures for Experiment Three

Each participant received a letter containing a description of the study and its purpose. The participants were provided with a link that contained a complete copy of the JAAT including the stimuli sheet and response booklet. Participants were asked to thoroughly review and examine the test using the time they deemed appropriate to do so. After consent was read, the participants were encouraged to use the link for the electronic online Qualtrics questionnaire and were informed that their initiation of the questionnaire is considered an electronic consent. A brief explanation of the concepts of cultural and linguistic relevance were introduced to all participants at the beginning of Questionnaire 2. Although Qualtrics estimated the range of time to complete Questionnaire 2 to be 15-20 minutes, the average completion time was actually 40 minutes and 26 seconds. This was probably due to SMEs taking their time to carefully review each subtest or part of the JAAT-3 before responding to questionnaire items.

For each subtest, four statements followed with seven response choices based on Likert scale: (7) *strongly agree*, (6) *agree*, (5) *somewhat agree*, (4) *neither agree nor disagree*, (3) *somewhat disagree*, (2) *disagree*, and (1) *strongly disagree*. The statements, shown below and in Appendix D, were concerned with the quality of the subtests of each part of the JAAT.

1. This subtest assesses what it is designed to assess.
2. This subtest is linguistically sensitive to the Jordanian population.

3. This subtest is culturally sensitive to the Jordanian population.
4. This was a culturally & linguistically sensitive subtest that provided enough information for me to adequately assess patient's expository speech.

The above items are specifically designed to address each subtest within a part of the JAAT-3. After a participant finished responding to all statements related to the subtests within a part, s/he was asked to judge whether the quantity of items presented in all subtests of the part were enough to assess what it was designed to assess. This questionnaire item addressed the four parts of the JAAT-3. It appraised the adequacy of the number of items presented within each part of the JAAT-3. The phrasing used in this question was as follows: "When considering all the three subtests of the Verbal Fluency part, does this part contain enough items to assess verbal fluency in Jordanian persons with aphasia?" In the case of the participant answering the question with "no" or "not sure," the participant was asked to elaborate more in a designated item as shown in Appendix E.

After the 48 subtests and four parts of the JAAT-3 were addressed in the questionnaire, the eight concluding items of the questionnaire assessed the entirety of the JAAT-3 were presented. The statements in this section of the questionnaire are shown in Appendix C.

Data Analysis for Experiment Three

Data was entered directly from Qualtrics Questionnaire 2 into SPSS data sheets for analysis. The SPSS data was stored on a password protected laptop. Box online storage provided by UNCG was used to store data securely. To address the research

questions related to the clinical applicability of the JAAT-3 as well as linguistic and cultural sensitivity, there were four series of data analyses run in this project:

- a. Series One examined *clinical applicability* of the JAAT-3 by computing means, medians, standard deviations, and interquartile ranges of all the questionnaire items appraising clinical applicability. These descriptive statistical analyses included all questionnaire items appraising clinical applicability on the JAAT-3 subtest level (Appendix D), as well as questionnaire items appraising the JAAT-3's four parts (Appendix E) and questionnaire items appraising the JAAT-3 as unit (questionnaire items 1, 3, 4, 5, 6, 7, and 8 in Appendix C).
- b. Series Two examined *linguistic relevance, cultural sensitivity and overall impression of SMEs* of the JAAT-3 by computing means, medians, standard deviations, and interquartile ranges of all the questionnaire items appraising linguistic relevance on the subtest level. Statistical analyses included one questionnaire item appraising the JAAT-3 linguistic and cultural sensitivity at the test unit level (Questionnaire 2 item 2 in Appendix C).
- c. Series Three: This series of statistical analysis is additional and does not pertain to a research question. It was primarily concerned with examining how the parts of the JAAT-3 performed in comparison to each other. Tabulations of means of scores on the four aspects of the JAAT-3 were formulated and compared on the subtests' levels. Additionally, percentile response rates on the four yes/no questionnaire questions asking about the adequacy of the

number of items in each part (Appendix E) were created for comparative purposes.

- d. Series Four: Cronbach's alpha, a measure of internal consistency that measures how closely related a set of items are as a group, was calculated for all subtest items as well as three of the items that assess the impressions of Jordanian SLPs on a grand test level. These items are unified by their use of a scale of seven responses making it possible to compute consistency (7 = "*strongly agree*" and 6 = "*agree*," etc.). This series of analyses examined the internal consistency of SME's responses on Questionnaire 2 in relation to their professional impressions on the JAAT-3. It is a dependable mean of computing the reliability of Questionnaire 2 as a scale for collecting information from SMEs.

Content Validity Index (CVI) was computed and documented as performed in other studies (Hughes, 1998; Hyrkäs, Appelqvist-Schmidlechner, & Oksa, 2003; Kyngas et al., 2000; Leung & Arthur, 2000; Lin, Chen, Chen, & Portwood, 2001). This type of validation is calculated by summing the percentage agreement scores of all items based on a previously set of criteria on all number of raters. Items that were given a rating of 7="strongly agree" and 6="agree" were summed on total number of raters: $CVI = \frac{\text{number of raters giving a rating of 6 or 7}}{\text{total number of raters}}$. In this study, items/subtests were considered adequate if there was an agreement of 80% or more. This analysis is imbedded with and presented in tables within Series One and Series Two analyses. CVI is computed only on questionnaire items with Likert scales of 1-7.

CHAPTER IV

RESULTS

Experiment One (Questionnaire 1: Verification of the JAAT-1 Visual Stimuli by NBD Jordanian Arabic Speakers)

Experiment One used a questionnaire to examine the clarity of the pictures in the first version of the JAAT, or JAAT-1. This was achieved by presenting the JAAT pictures to 164 NBD Jordanian Arabic speaking participants. Table 2 presents the numbers of words distributed by the percentages of correct responses by all participants. One hundred twenty-one words (97.5% of all pictures to be used in the JAAT) had 90% or more correct response rate. This criterion was created for this study by the author of the study and the advisor. There were three pictures that had less than the 90% criterion of correct identification: *Tie* (89% correct response rate), *screw* (87% correct response rate), and *lighter* (60% correct response rate). Ten pictures were repeated to assess consistency of responses; they showed no marked difference in performance within participants. All repeated pictures had a more than 90% correct response rate. As expected, there was a high success rate within participants in identifying the pictures with no significant differences based on age or educational background. The picture of a *Tie* was not included in the JAAT-1, the picture of a *screw* was deleted and replaced by the picture of a *key*, and an alternative to the *boy with the lighter* picture was created to be used in Session 2 of Experiment Two. At this stage, JAAT-2, a lightly modified version of the JAAT-1 was created for use in Experiment Two.

Table 2

Number of Words Distributed by Percentages of Responses on Questionnaire 1

Percentages of agreement	Number of words
100%	58
95% to 99%	57
90% to 94%	3
Less than 90%	3
Total number of words	121

Experiment Two (Identifying Inadequate Items in the JAAT-2 by Administering the Test on Senior NBD Jordanian Arabic Speakers)

The purpose of Experiment Two was to identify any outlier items in the JAAT-2, or items that 30% or more of the examinees responded incorrectly to. Any items that met the deletion criteria were marked for modification or replacement. It is safe to assume that any item found by 30% of a relatively homogeneous group to be challenging would be even more challenging for PWA. This criterion was created for the study by the author of the study and advisor. Based on extensive review and search by the author of the present study, no similar methodology has been used in aphasia testing prior to this attempt. To identify item outliers in the JAAT-2, the test was administered to 20 NBD Jordanian Arabic-speaking individuals living in three cities in North Carolina. The average age was 68 years and 7 months with the minimum age being 60 and the maximum age being 80. Eleven participants were females and nine were males. Educational status varied from illiterate to graduate degree holders. Four participants were illiterate, two participants had 1-6 years of school education, seven participants had

1-12 years of school education, five participants had undergraduate degrees, and two participants had graduate degrees. Participants reported no history of cerebrovascular incidents (CVAs), dementia, or any history of brain injury. While most participants used eye glasses, none of them reported any visual impairments that made seeing the test items difficult. No subject used hearing aids or reported hearing difficulties.

Table 3 presents the overall performance of participants on the JAAT subtests and parts as well as the total scores. The average score on the JAAT-2 was 104.75 from a potential maximum score of 112. Nine out of 112 items used in the JAAT had repeated incorrect responses of 30% or more. The three subtests that had items with less than 71% success rate were in the Verbal Fluency and Naming parts. The subtests in the Verbal Fluency part were *picture description* and *picture description of sequences* addressing expository speech.

Table 3

Average Performances of Subjects on the JAAT-2

Parts and Subtests	Minimum score	Maximum score	<i>M</i>	<i>SD</i>
Verbal Fluency				
<i>Open ended questions (10)*</i>	10	10	10	0
<i>Picture description (6)</i>	2	6	4.45	1.23
<i>Picture sequence description (6)</i>	4	6	5.4	0.68
Overall Part Score (22)	16	20	19.85	1.75
Auditory Comprehension				
<i>Following commands (9)</i>	6	9	8.35	0.81
<i>Single word comprehension (20)</i>	17	20	19.40	1.04
<i>Sentence comprehension (6)</i>	4	6	5.55	0.63
<i>Sentence & paragraph comprehension (12)</i>	6	12	10.95	1.63
Overall Part Score (47)	35	47	44.25	3.56

Table 3

Cont.

Parts and Subtests	Minimum score	Maximum score	<i>M</i>	<i>SD</i>
Repetition				
<i>Repetition of words non-words and sentences (18)</i>	14	18	17.5	0.99
Overall Part Score (18)	14	18	17.5	0.99
Naming				
<i>Object naming (5)</i>	5	5	5	0
<i>Picture naming (15)</i>	10	15	13.55	1.63
<i>Responsive naming (5)</i>	4	5	4.9	0.3
<i>Categorical naming**</i>	11	26	17.5	4.11
Overall Part Score (25)	19	25	23.45	1.79
Overall Test Score (112)	84	111	104.75	7.35
Time (minutes: seconds)	17:5	30:53	23:04	3.91

Note. * Possible maximum score on subtest and part; ** Categorical naming varies depending on individuals

In the naming part, the subtest of *picture naming* had four items that proved to be difficult for 30% or more of participants. On the “Guest Room” picture in the *picture description* subtest within the Verbal Fluency part, nine of the 20 subjects (45% of total number of sample) were not able to describe the picture of the cat’s tail being stepped on by the girl. Nine subjects did not identify the action of the girl stepping on the cat’s tail and five subjects described the cat as a dog or described it as an animal. Seven subjects (55% of total number of sample) neither identified the action of the man reading a book, nor did they identify the man’s traditional clothing.

In the *Guest Room* picture description subtest, the picture of a *young boy playing with a lighter* was designed to elicit verbal responses from examinees by provoking attention to an alarming event. Eleven subjects (55% of total number of sample) were not able to identify the object that the child in the picture held in his hand (*lighter*). This was

a picture that proved to be challenging for participants in Experiment One with 40% incorrect responses. Finally, four pictures had less than 71% correct responses in the Naming Part's picture naming subtest. These were pictures of a *carpet* and a *stapler* and actions of *ironing*, and a *biting dog*. Six out of 20 subjects (30% of the group) were unable to successfully name each picture.

The picture of the *boy with the lighter* failed to pass the criterion of agreement in Experiment One and Experiment Two and proved to be challenging to correctly identify for both groups. This means that the picture failed to achieve its purpose which required an alternative picture to be designed. A picture of a *boy opening a bird's cage* leading to a parrot to escaping was used in Session 2 of Experiment One and achieved the inclusion criterion for the final version of the JAAT. The picture of the girl stepping on the cat's tail was also graphically modified to make the action easier to identify by improving the depiction of the animal and creating some animation highlights surrounding the objects.

The depiction in the Verbal Fluency subtest of *the man with the traditional attire* was removed due to lack of recognition and was found to be superfluous. Appendix G shows the alternative pictures that were used in Session 2 of Experiment Two to be integrated in the final version of the JAAT. At this stage of modification, a final version of the JAAT was ready to be used in Experiment Three. The four pictures in the Naming part were replaced by the pictures of a *cigarette* and a pair of *scissors* and pictures of a *man sweeping* and a *man milking* a cow. These pictures were presented to participants in Experiment One and achieved more than 90% correct responses.

Finally, six participants (30% of the group) did not name two objects on the table in subtest *Picture description of sequence* in the Verbal Fluency part of the JAAT-2. Previous pictures were replaced by alternates in this subtest to improve the quality of this item as shown in Appendix G part 2. The results indicate that nine out of the 112 items used in the JAAT-2 required deletion or modification. This number represents 8% of the overall number of items in the JAAT-2. Table 4 shows the list of items that required cancelation and/or deletion. At this stage, and after the necessary modification of items described above were made on the JAAT-2, the JAAT-3 was produced and ready to be appraised by SMEs in Experiment Three.

Table 4

Items with 30% or More Incorrect Responses on the JAAT-2

Parts and Subtests	Number of incorrect responses	Percentage of incorrect response from group
Verbal Fluency Part		
Picture description subtest:		
<i>Cat identified as cat</i>	9	45%
<i>Cat tail being stepped on by girl</i>	9	45%
<i>Man wearing traditional attire reading book</i>	7	35%
<i>Lighter held in boy's hand</i>	11	55%
Picture sequence description Subtest:		
<i>Naming two objects on kitchen table</i>	6	30%
Naming Part		
Picture naming subtest:		
<i>Carpet</i>	6	30%
<i>Stapler</i>	7	35%
<i>Biting</i>	6	30%
<i>Ironing</i>	6	30%

Experiment Three (Questionnaire 2: Validation of Test Content Based on SME Rating of the JAAT-3)

The purpose of this experiment was to assess the new revised JAAT-3's content based on the three parameters of clinical applicability, linguistic relevance, and cultural sensitivity to the Jordanian PWA population. To achieve that purpose, an electronic copy of the JAAT-3 was sent to 21 SMEs who were asked to respond online to Questionnaire 2 (validation of test content based on SME rating of the JAAT). Twenty-one certified Jordanian SLPs with more than 2 years of experience participated in Experiment Three as SMEs to appraise the JAAT-3. Academic credentials of the participants ranged from undergraduate (*four participants with Bachelor's degree in SLP*) to graduate (*13 participants with Master's degrees in SLP and four participants with Ph.D. degrees in SLP*). Years of clinical experience with PWA ranged from four participants with 2-5 years (*19% of total number of participants*), twelve with 5-10 years (*57% of total number of participants*), and five with 10 years or more (*23.8% of total number of participants*).

Jordanian official clinical credentialing allows Master's degree or Ph.D. holders in speech-language pathology to work independently as SLPs after passing a clinical certification exam. Bachelor's degree holders of speech-language pathology can practice under indirect supervision—from certified Master's degree holding SLPs—after passing a certification exam. As well as holding local Jordanian clinical certification, all the Ph.D. holders and three Master's degree holders had American clinical certification. Three Master's degree holders had other unspecified clinical certification which may indicate that they were working outside Jordan. No information was obtained on location of current places of practice.

Only participants who completed answering all the questionnaire's items were included in the study—10 eligible participants did not complete the whole questionnaire and so were excluded from the study. Twenty-one participants responded to all the items in the questionnaire.

As previously explained, online Qualtrics Questionnaire 2 (validation of test content based on SME rating of the JAAT) assessed the subtests of the JAAT-3, its parts, and the JAAT-3 as a whole test. Subtests were addressed using statement items shown in Appendix D. These questionnaire items examined three major aspects: (a) clinical applicability, (b) linguistic relevance, (c) cultural sensitivity, as well as (d) overall impression. Next, each of the four parts of the JAAT-3 were addressed with one yes/no question (Appendix E). Then, eight questionnaire items were presented to appraise the JAAT-3 as a whole test. All the items in the questionnaire were statistically analyzed according to the above order in this section. Finally, an additional analysis was conducted; Cronbach alpha reliability analysis was calculated to measure the internal consistency of Questionnaires 2 (validation of test content based on SME rating of the JAAT-3). These analyses are presented below in four analysis series as previously explained in the methods section of the study.

Series One Statistical Analysis: Examining the Clinical Applicability of the JAAT-3

Comprehensive aphasia tests are clinical tools that can be applied to collect information for clinical use by covering multiple aspects of communication necessary for diagnosis, prognosis, and treatment (Ivanova & Hallowell, 2013). For a test to be considered clinically applicable, it must exhibit traits that are valued by clinicians such as

having a relatively short administration time and capable of collecting information that can serve diagnostic purposes (Marshall & Wright, 2007). This series of analysis examined SMEs professional impressions of the JAAT-3 on the subtest level as well as parts level. Additionally, the questionnaire items pertaining to the test as a whole was also analyzed. CVI analysis results were included in this series.

The first set of analyses in this series examined questionnaire items on the subtest level by calculating means (*M*), medians, standard deviations (*SD*), and interquartile ranges (*IR*) for each item. Additionally, Content validity index (*CVI*) was calculated based on SME ratings. For questionnaire items assessing “clinical applicability” of the JAAT, Table 5 presents descriptive statistics for the questionnaire items that assessed the clinical applicability of the JAAT along with content validity indices. A Likert scale was used as follows: 7=*strongly agree*, 6=*agree*, 5=*somewhat agree*, 4=*neither agree nor disagree*, 3=*somewhat disagree*, 2=*disagree*, and 1=*strongly disagree*.

Table 5

Questionnaire Items Appraising Clinical Applicability of Subtests to Test Aphasia in Jordanian Arabic Speakers

Parts and Subtests	<i>M</i>	Median	<i>SD</i>	IR	CVI
Verbal Fluency					
<i>social responses Subtest linguistically sensitive to Jordanians</i>	6.61	7	.80	0	81%
<i>Subtest can assess expository speech</i>	6.42	7	.97	1.5	76%
<i>Subtest can assess sequenced expository speech</i>	6.52	7	.51	1	100%

Table 5

Cont.

Parts and Subtests	<i>M</i>	Median	<i>SD</i>	IR	CVI
Auditory Comprehension					
<i>Subtest can assess comprehension of simple commands</i>	6.57	7	.50	1	100%
<i>Subtest can assess comprehension of Single words</i>	6.71	7	.46	1	100%
<i>Subtest can assess comprehension of sentences</i>	6.66	7	.48	1	100%
<i>Subtest can assess comprehension of sentences & paragraph</i>	6.38	6	.74	1	95%
Repetition					
<i>Subtest can assess repetition of words non-words and sentences</i>	6.61	7	.73	1	95%
Naming					
<i>Subtest can assess object naming</i>	6.57	7	.50	1	100%
<i>Subtest can assess picture naming</i>	6.76	7	.43	.5	100%
<i>Subtest can assess responsive naming</i>	6.38	7	.92	1	90%
<i>Subtest can assess categorical naming</i>	6.66	7	.73	.5	95%

Note. *M* for all means in questionnaire items assessing clinical applicability: 6.56

As expected, most subtests were found by SMEs to exhibit strong potential of clinical applicability for Jordanian PWA as evidenced by high means and medians. Conformity of responses were also indicated by low levels of dispersion. CVI criterion rate was set to 80% as with other studies (Hughes, 1998; Hyrkäs et al., 2003; Kyngas et al., 2000; Leung & Arthur, 2000; Lin et al., 2001). As can be seen in Figure 1, most participants responded positively with high agreement on all subtests (CVI rates were all above the criterion of 80%) with the only exception being subtest 2 assessing expository speech (CVI rate 76%) which resulted in an overall lower CVI. This means that this subtest containing the *Guest Room* picture required more improvement to make it better

suites to elicit verbal responses from future examinees. Finally, all means were added and divided by their total number to create a value representing the mean of the means. Overall, the average mean on all subtests appraising the clinical applicability of the JAAT was 6.56 with average CVI of 94%. The descriptive statistic results indicate a positive high agreement rate overall based on SMEs professional judgment of the subtests in the JAAT-3. SMEs high average rating of 6.56 represents a favorable indication for the future applicability of the JAAT-3 in clinical use.

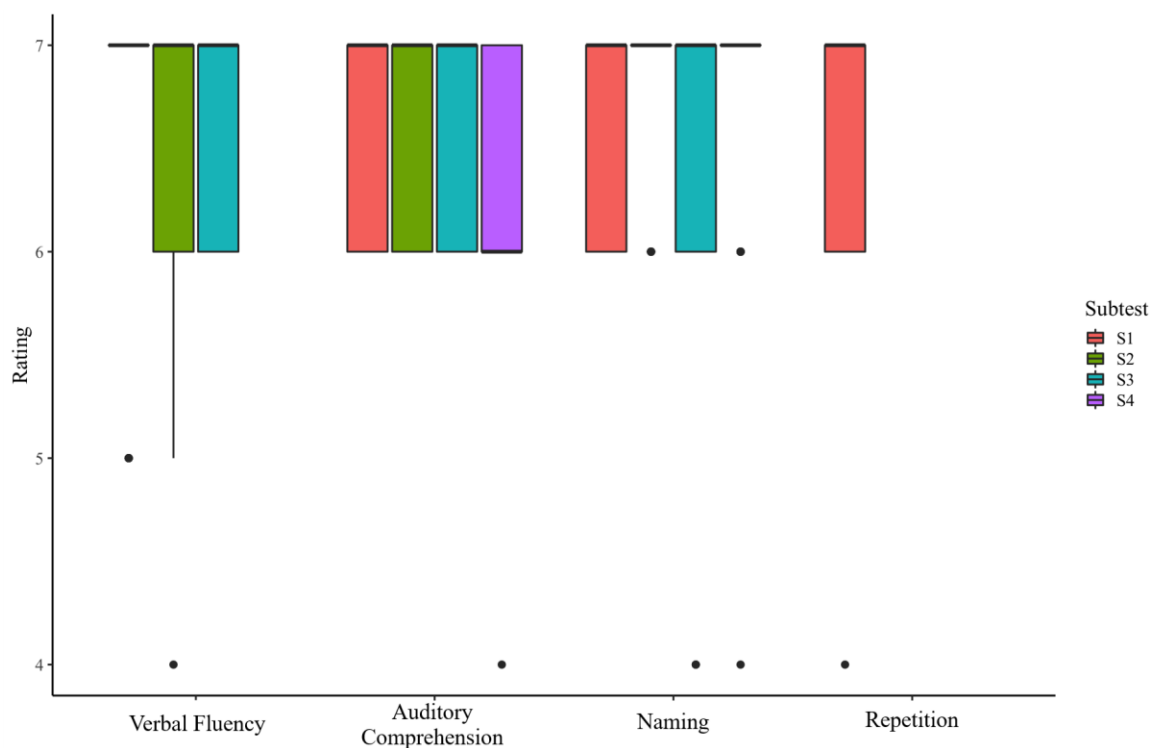


Figure 1. Data Distribution of the 21 SMEs Responses on Clinical Applicability of Subtests in the JAAT-3. Each Rectangle Represents a Subtest within the Designated Part of the Test.

CVI rating was set purposely high (80%) as previously explained and the resulting high average rating exceeded the set standard indicating advantageous content validity of the JAAT-3 as far as clinical applicability is concerned. The results mean that more than 80% of SMEs responded positively by choosing 7 “*Strongly agree*” or 6 “*Agree*” on clinical applicability on almost all subtests. The exception was for *expository speech* assessing subtest in the Verbal Fluency part of the JAAT-3. This subtest of the JAAT-3 has room for improvement as descriptive statistics and CVI value have indicated.

While subtest questions target the quality of subtests in the JAAT-3, part-specific questions reflect clinical importance through collecting judgment of SMEs that pertains to whether the number of items in each part is adequate in assessing the language ability the part is targeting (Appendix E). This is the second set of analysis in Series One where the results of the four yes/no questions addressing each of the four parts of the JAAT-3 are calculated. Table 6 presents the results of the 21 SMEs’ impressions regarding each part of the JAAT-3 and shows the percentages of agreement for each question. Results show agreements higher than 80% which indicate that most SMEs viewed all four parts of the JAAT-3 to contain adequate numbers of items to assess verbal fluency, auditory comprehension, repetition, and naming.

The third set of statistical analyses in this series is the analysis of seven Questionnaire 2 items that evaluate the JAAT-3 as a whole test (Appendices C & F). Seven out of eight questions (Table 7) appraise the clinical values of the JAAT-3 the

exception to this being the second question that asked the SMEs to give their professional impression on whether the JAAT-3 is linguistically and culturally sensitive or not.

Table 6

Results Showing Agreement as to Whether Each Part Contains Enough Items in its Subtest to Assess What it Was Designed to Assess

Item in Questionnaire	Response options		
	Yes	No	Not Sure
Verbal Fluency			
<i>Number of participants</i>	18	1	2
<i>Percentage</i>	85.7%	4.8	9.5
Auditory Comprehension			
<i>Number of participants</i>	19	1	1
<i>Percentage</i>	90.5%	4.8%	4.8%
Repetition			
<i>Number of participants</i>	20	0	1
<i>Percentage</i>	95.2%	0%	4.8%
Naming			
<i>Number of participants</i>	20	0	1
<i>Percentage</i>	95.2%	0%	4.8%

The expected time that it may take to administer the JAAT-3 was judged by SMEs to be manageable. Twenty participants (95%) answered “yes” when asked whether the JAAT-3 was suitably timed or not, with only one participant reporting that the test was too long. When asked how likely it was that a participant would use the JAAT-3 if it was available, 18 participants (86%) answered with “likely,” two participants (9.5%) answered with “somewhat likely,” and one participant answered with “not sure.” Seventeen SMEs (81%) responded that the JAAT-3 would be their first choice if it was available, and four SMEs (19%) responded that the JAAT-3 would be their second

choice. These results reinforce the overall flow of positive impressions the JAAT-3 has generated from the 21 SMEs participating in the study.

Table 7

Questionnaire Items Appraising the JAAT-3 as a Testing Unit

Questions	Type and quantity of responses				
	Too Long	Too Short	Suitably Timed		
Q1. <i>How would you describe administration time of the test?</i>	1 (4.8%)	0 (0.0%)	20 (95%)		
	Yes	No	Not Sure		
Q2. <i>Overall, do you consider this test linguistically and culturally sensitive for Jordanian persons with aphasia?</i>	19 (90.5%)	0 (0.0%)	2 (9.5%)		
Q6. <i>Would this test be useful in identifying other impairments associated with aphasia such as apraxia and dysarthria?</i>	0 (0.0%)	8 (38%)	13 (62%)		
	Likely	Somewhat Likely	Not Sure	Somewhat Unlikely	Unlikely
Q7. <i>If this test was available to you for use right now, how likely would you be to use it?</i>	18 (85.7%)	2 (9.5%)	1 (4.8%)	0 (0.0%)	0 (0.0%)
	First Choice	Second Choice	Not Sure	Will Not Use	
Q8. <i>If this test is available for you now, how would you rank it based on potential use</i>	17 (81%)	4 (19%)	0 (0.0%)	0 (0.0%)	

The remaining four questionnaire items appraising the JAAT-3 as a whole test are important clinically. When asked if the JAAT-3 can assess disorders other than aphasia such as dysarthria and apraxia, 13 SMEs (62%) answered “no” and eight (38%) answered “not sure.” This result is a significant indication that the JAAT-3 cannot generalize its use to other disorders such as apraxia and dysarthria if it remains unchanged. Three SMEs specifically mentioned the lack of items examining apraxia of speech making it clear that apraxia cannot be identified with the current version of the JAAT-3.

Table 8 presents the results of the last three questionnaire items that address three important aspects of any clinical assessment tool: the capacity of the JAAT-3 to (a) provide a diagnostic description of receptive and expressive language in PWA, (b) determine severity level in PWA, and (c) help in designing an appropriate treatment plan for PWA. Table 9 shows how raters reported strong agreement on the potential use of the JAAT-3 to gather diagnostic and treatment information as indicated by consistently high means and medians and greater than 80% CVIs. The use of the JAAT-3 to determine severity levels in Jordanian Arabic-speaking PWA is questionable as the descriptive statistical values are low (Figure 2) and CVI value was below the criterion of 80%.

Table 8

Questionnaire Items Appraising the JAAT-3 as a Testing Unit

Questions	Type and quantity of responses*				
	Strongly Agree	Agree	Somewhat agree	Undecided	Disagree
Q3. <i>The test can provide a diagnostic description of receptive and expressive language in persons with aphasia</i>	11 (52%)	6 (28.6%)	4 (19%)	0 (0.0%)	0 (0.0%)
Q4. <i>The test can determine the severity levels within persons with aphasia</i>	5 (23.8%)	9 (42.9%)	6 (28.6%)	1 (4.8%)	0 (0.0%)
Q5. <i>The test can help in designing a treatment plan for persons with aphasia</i>	6 (28.6%)	12 (57.1%)	3 (14.3%)	0 (0.0%)	0 (0.0%)

*There are seven options in Questionnaire 2. They were not included in this table due to no responses to "Somewhat Disagree" and "Strongly Disagree" were reported.

Table 9

Questionnaire Items Appraising Important Clinical Aspect of the JAAT-3 as a Unit

Questionnaire item	<i>M</i>	Median	<i>SD</i>	IR	CVI
<i>Test can provide a diagnostic description of receptive and expressive language in persons with aphasia</i>	6.33	7	.79	1	81%
<i>Test can determine severity levels within persons with aphasia</i>	5.86	6	.85	1.5	67%
<i>Test can help in designing a treatment plan for persons with aphasia</i>	6.14	6	.65	1	85%

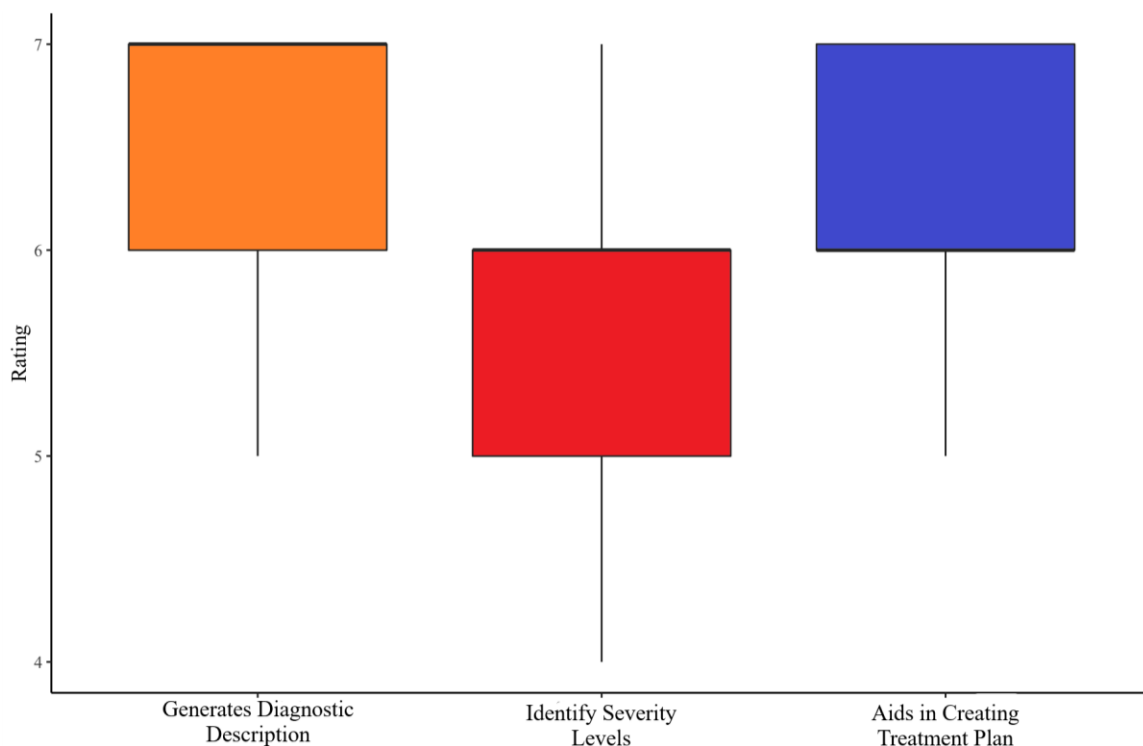


Figure 2. Range Distribution of Response Scores on Items Assessing Clinical Attributes of the JAAT-3.

Series Two Statistical Analysis: Examining Linguistic and Cultural Sensitivity and Overall Impressions of the JAAT-3

Linguistic relevance or sensitivity is the quality of agreement and appropriateness of linguistic content to sociolinguistic aspects of a certain group (Ball, Perkins, Müller, & Howard, 2008). Linguistic relevance requires sensitivity to what words are used and how they fit within the language of the target population (Gales, 2003).

Means, medians, standard deviations, interquartile ranges, and content validity indices of the questionnaire items appraising the “linguistic relevance” of subtests in the JAAT-3 were computed as presented in Table 10. Raters responded positively on most subtests as seen in the high means and medians. They also appear to be consistent due to

the low values of IR range and standard deviation indicating low dispersion. On all the subtests appraising the linguistic relevance of the JAAT, the average mean was 6.49 with an average CVI of 92%. SMEs high average rating of 6.49 indicate high linguistic sensitivity of the JAAT-3. High average rating on CVI proves high content validity of the JAAT-3 on the subtest level. This means high linguistic relevance in the JAAT-3 on almost all subtests.

Table 10

Questionnaire Items Appraising “Linguistic Relevance” of Subtests to Jordanian Arabic Speakers

Parts and Subtests	<i>M</i>	Median	<i>SD</i>	IR	CVI
Verbal Fluency					
<i>Subtest on simple social responses</i>	6.52	7	.81	1	81%
<i>Subtest on expository speech</i>	6.09	7	1.13	2	62%
<i>Subtest on sequenced expository speech</i>	6.42	6	.50	1	100%
Auditory Comprehension					
<i>Subtest on comprehension of simple commands</i>	6.57	7	.50	1	100%
<i>Subtest on comprehension of Single words</i>	6.61	7	.49	1	100%
<i>Subtest on comprehension of sentences</i>	6.42	6	.50	1	100%
<i>Subtest on comprehension of sentences & paragraph</i>	5.95	7	1.32	3	71%
Repetition					
<i>Subtest on repetition of words non-words and sentences</i>	6.57	7	.74	1	95%
Naming					
<i>Subtest on object naming</i>	6.76	7	.43	0.5	100
<i>Subtest on picture naming</i>	6.61	7	.74	1	95%
<i>Subtest can assess responsive naming</i>	6.52	7	.74	1	95%
<i>Subtest can assess categorical naming</i>	6.80	7	.40	0	100%

Note. M for all means in questionnaire items assessing linguistic relevance: 6.49.

As in the clinical subtest aspect analyzed above, Subtest 2 continued to perform below the content validity criterion with only 62% CVI. Another subtest with lower mean and CVI values was Subtest 7 that assessed *comprehension of spoken sentences and paragraphs* using yes/no questions after a short statement and two short stories. These two subtests of the JAAT-3 showed room for improvement as descriptive statistics and CVI value have indicated. Figure 3 box plots show the skewed distribution of the two subtests along with borderline performance of Subtest 1.

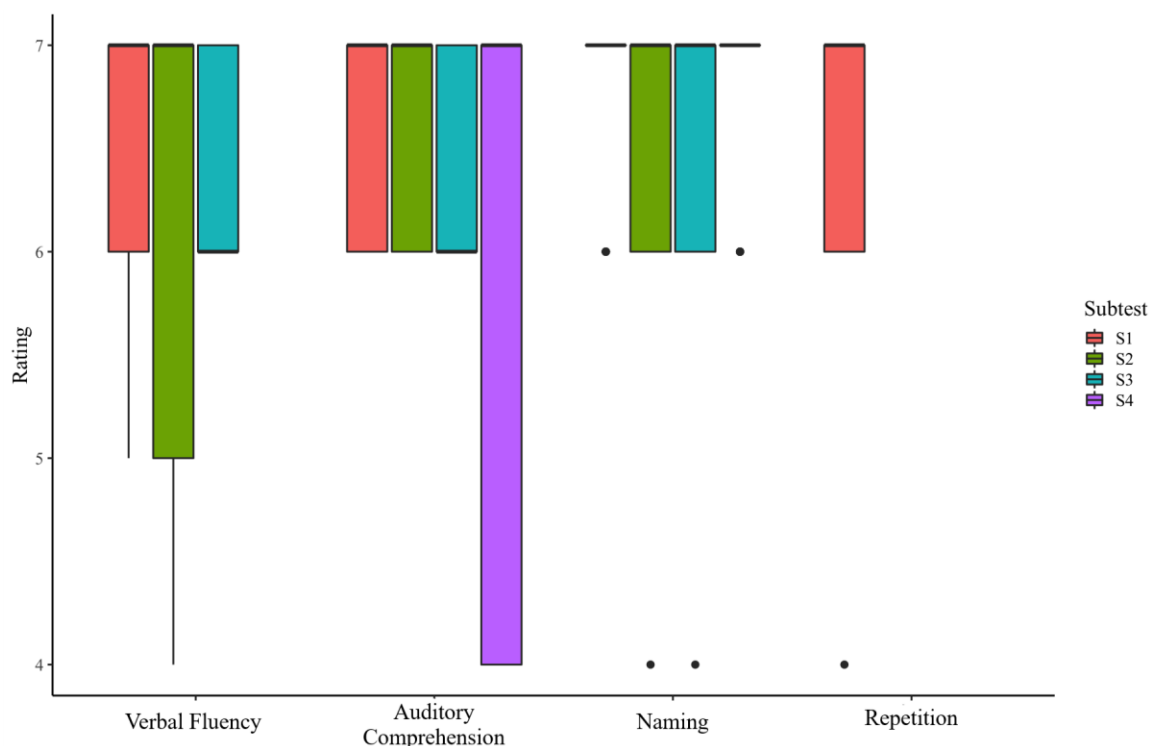


Figure 3. Data Distribution of the 21 SMEs Responses on Linguistic Relevance of Subtests in the JAAT-3. Each Rectangle Represents a Subtest within the Designated Part of the Test.

It is very important for an aphasia test to take into account the distinctive language backgrounds of a group and its acceptable mainstream cultural traits. Each

society has a diverse population yet there are usually common cultural features with a mutual set of values and cultural traditions. On the content level, assessment tools must be guided by and conform to the culture of the target population (del Rosario Basterra, Trumbull, Solano-Flores, & Solano Flores, 2011).

Subtests assessing cultural sensitivity were analyzed and are presented in Table 11. The overall response of raters showed significant positive impressions evidenced by high means and medians with strong consistency based on low standard deviations and interquartile ranges as expected. With clinical applicability and linguistic relevance subtests, Subtest 2 in the verbal fluency part showed a lower mean, higher standard deviation, and a low CVI of 76%. Figure 4 visually represents Subtest 2 as an outlier when compared to the subtests in the other four parts of the JAAT-3. Overall, the cultural sensitivity subtests had an average mean of 6.51 and average CVI of 95%. These favorable findings mean that the JAAT-3 is highly suitable for the Jordanian population.

These findings are corroborated at the level of appraisal of the JAAT-3 as a whole test. The second question in the set of eight questions in Questionnaire 2 assessing the JAAT-3 as a whole test asked the SMEs to give their professional impression as to whether the JAAT-3 is linguistically and culturally sensitive. As expected, 19 SMEs (90.5% of participants) reported that the JAAT was both linguistically and culturally sensitive to the Jordanian Arabic speaking population. Only two SMEs responded with “not sure.” This result is consistent with the overall high average means on linguistic and culturally specific subtest ranging from 6.49 to 6.56. This also corroborates the overall average CVI of subtests targeting linguistic and cultural sensitivity. Overall, the SMEs

perceived the JAAT-3 to be a linguistically and culturally sensitive test for the Jordanian Arabic-speaking population.

Table 11

Questionnaire Items Appraising “Cultural Sensitivity” of Subtests to Jordanian Arabic Speakers

Parts and Subtests	<i>M</i>	Median	<i>SD</i>	IR	CVI
Verbal Fluency					
<i>Subtest on simple social responses</i>	6.80	7	.60	0	100%
<i>Subtest on expository speech</i>	5.80	7	1.32	2.5	76%
<i>Subtest on sequenced expository speech</i>	6.38	6	.49	1	100%
Auditory Comprehension					
<i>Subtest on comprehension of simple commands</i>	6.57	7	.50	1	100%
<i>Subtest on comprehension of Single words</i>	6.57	7	.50	1	100%
<i>Subtest on comprehension of sentences</i>	6.42	6	.50	1	100%
<i>Subtest on comprehension of sentences & paragraph</i>	6.28	7	1.05	1	85%
Repetition					
<i>Subtest on repetition of words non-words and sentences</i>	6.47	7	1.07	0.5	86%
Naming					
<i>Subtest on object naming</i>	6.66	7	.73	0.5	95%
<i>Subtest on picture naming</i>	6.66	7	.73	0.5	95%
<i>Subtest can assess responsive naming</i>	6.66	7	.48	1	100%
<i>Subtest can assess categorical naming</i>	6.80	7	.40	0	100%

Note. M for all means in questionnaire items assessing cultural sensitivity: 6.51

mean scores on the subtests assessing overall impression of raters. Average mean of subtest appraising the overall impression of SMEs was 6.50 and a CVI average of 94%. At this point in the analysis, raters viewed most subtests of the JAAT-3 to show potential for clinical use with the exception of Subtest 2. For linguistic relevance, raters viewed most subtests to be suitable for the language of Jordanian Arabic speakers with the exclusion of Subtest 2 and 7. Finally, the cultural sensitivity of most subtests in the JAAT-3 was strong except for Subtest 2. These findings indicate that the JAAT-3 scored very high overall with SMEs. This means that the test's content represented high potential for clinical applicability and showed high linguistic and cultural sensitivity to the Jordanian population.

Table 12

Questionnaire Items Appraising the "Overall" Impression of SME on Subtests

Parts and Subtests	<i>M</i>	Median	<i>SD</i>	IR	CVI
Verbal Fluency					
<i>Subtest on simple social responses</i>	6.38	7	1.02	2	95%
<i>Subtest on expository speech</i>	6.04	7	1.16	2	90%
<i>Subtest on sequenced expository speech</i>	6.38	6	.49	1	100%
Auditory Comprehension					
<i>Subtest on comprehension of simple commands</i>	6.52	7	.51	1	100%
<i>Subtest on comprehension of Single words</i>	6.71	7	.46	1	100%
<i>Subtest on comprehension of sentences</i>	6.57	7	.50	1	100%
<i>Subtest on comprehension of sentences & paragraph</i>	6.23	7	1.04	1	85%

Table 12

Cont.

Parts and Subtests	<i>M</i>	Median	<i>SD</i>	IR	CVI
Repetition					
<i>Subtest on repetition of words non-words and sentences</i>	6.57	7	.92	0.5	90%
Naming					
<i>Subtest on object naming</i>	6.52	7	.74	1	90%
<i>Subtest on picture naming</i>	6.76	7	.43	0.5	100%
<i>Subtest can assess responsive naming</i>	6.57	7	.74	1	95%
<i>Subtest can assess categorical naming</i>	6.71	7	.71	0	95%

Note. M for all means in questionnaire items assessing linguistic relevance: 6.50

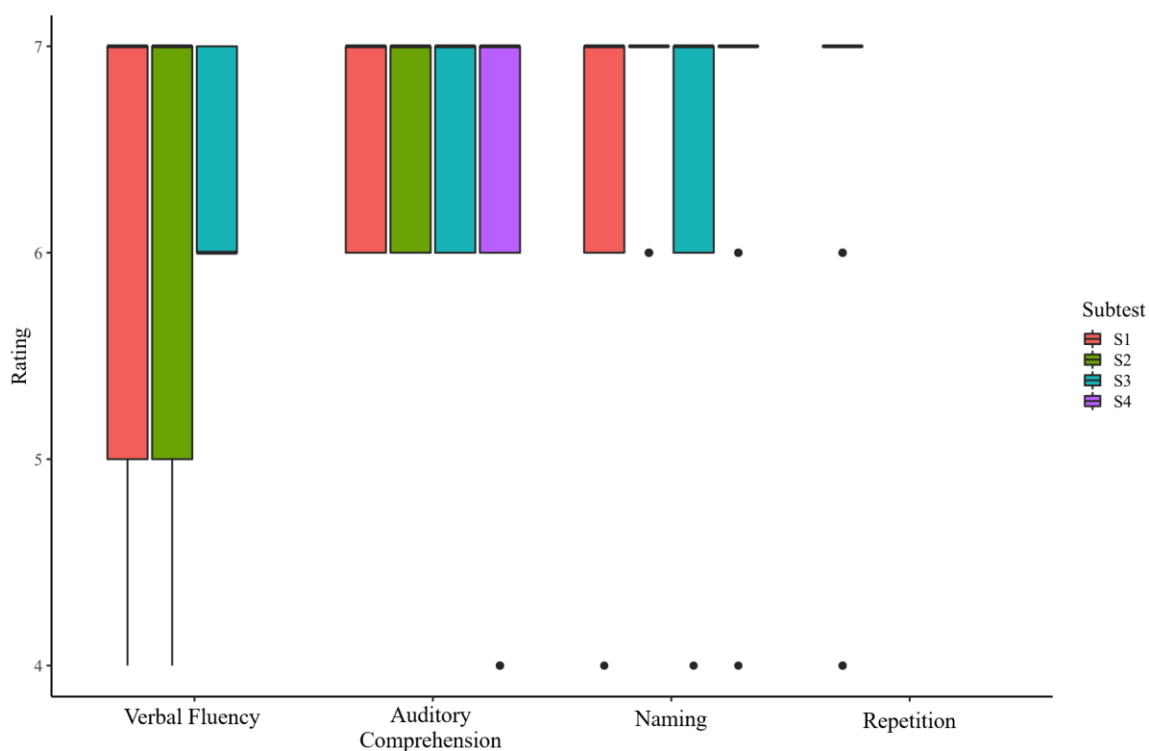


Figure 5. Data Distribution of the 21 SMEs Responses on Overall Impressions of Subtests in the JAAT-3. Each Rectangle Represents a Subtest within the Designated Part of the Test.

Series Three Statistical Analysis: Comparing the Performance of the Four Parts of the JAAT-3 to Each Other

Another way of looking at the data is by examining how the parts of the JAAT-3 performed compared to each other. This series of analyses is additional and does not address any of the research questions of this study. The purpose here is to weigh each part of the JAAT-3 in comparison to the others.

Average means on each of the four parameters of appraisal in Questionnaire 2 (clinical applicability, linguistic relevance, cultural sensitivity, and overall impression) were computed for each part of the JAAT-3. Figure 6 tabulates the means of the scores on the four aspects of appraisal to which each part was subjected. For example, in the Verbal Fluency part shown in the figure, the average response on *overall impression* (illustrated in yellow columns) was 6.25 while it was 6.65 on the Naming part of the JAAT-3. Those averages are both high considering that the maximum is 7 “Strongly agree” but they show how the Naming part of the JAAT-3 had better responses from SMEs.

The figure highlights the overall positive impression the raters had on the JAAT-3’s subtests and parts. The Naming part of the JAAT-3 had better ratings than all other parts and is followed consecutively by the Repetition part, the Auditory Comprehension part, with Subtest 7 showing less than the desired CVI in linguistic relevance, and finally, the Verbal Fluency part that contained Subtest 2, the least performing subtest in the JAAT-3 which scored lower than acceptable on clinical applicability, linguistic relevance, and cultural sensitivity.

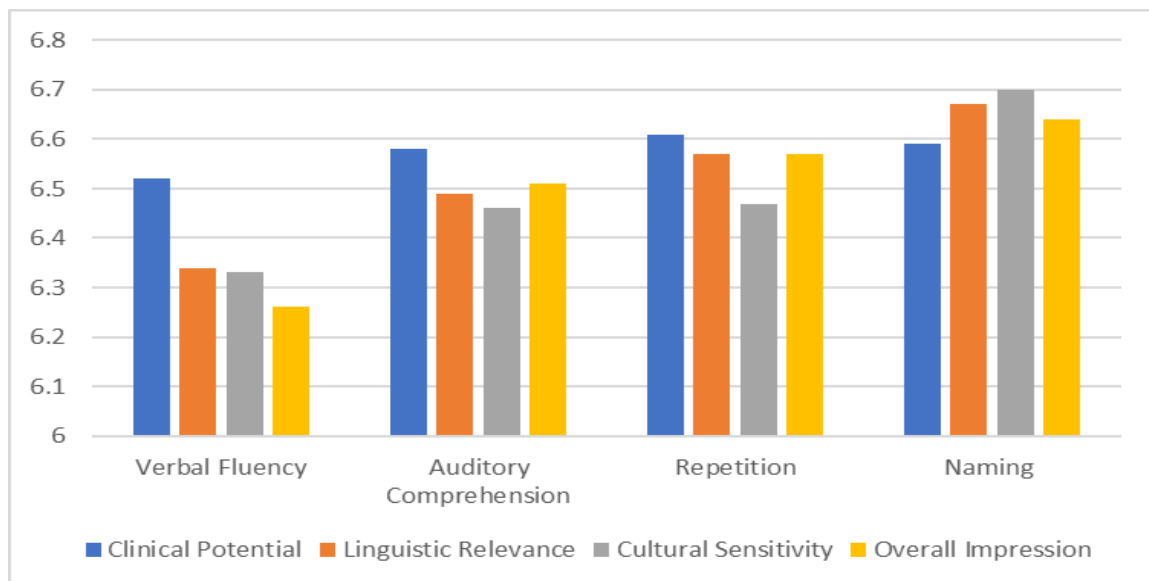


Figure 6. Mean Response Scores on Subtests within the Four Parts of the JAAT-3.

In Figure 6, all the averages of the parameters of appraisal in Questionnaire 2 (clinical applicability, linguistic relevance, cultural sensitivity, and overall impression) were collectively computed and represented in the blue columns within the figure. Each column represented the overall performance of each of the four parts of the JAAT-3. The difference between Figure 6 and Figure 7 is that in Figure 7 the four columns from Figure 6 were combined statistically in blue columns. These blue columns are then compared with the percentile agreement rates on the four questions in Questionnaire 2 found in Table 8. These percentiles reflect similar trends within SMEs when it comes to the parts of the JAAT-3. For example, 85% of SMEs found that the number of items available in the JAAT-3 is sufficient to assess Verbal Fluency. While this is a high value, 94% of the SMEs responded that the Naming part of the JAAT-3 had enough items to assess naming. These percentages show that although all parts of the JAAT-3 had high positive

impressions from SMEs, some were higher than others. These findings also indicate areas where more work is needed in the JAAT-3.

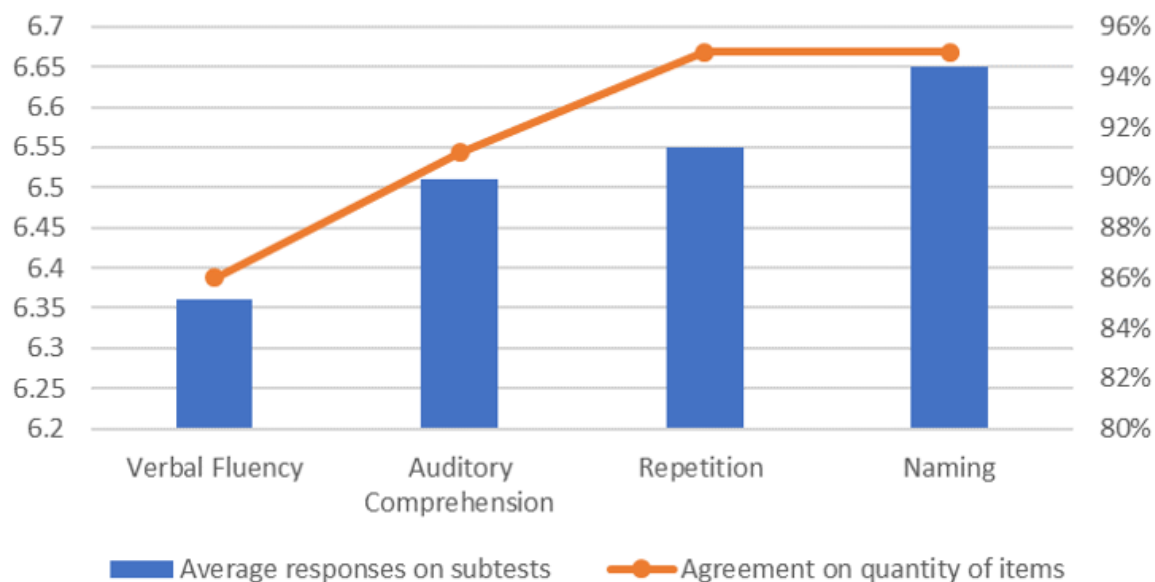


Figure 7. Mean Response Subtests' Scores of Parts with Percentile Agreement on Parts Questions.

Series Four Statistical Analysis: Examining the Consistency of SME Responses on Questionnaire 2

Experiment Three relies on SMEs' responses from Questionnaire 2; therefore, it is crucial to examine the consistency in these responses to make sure that the outcome of this questionnaire is a reliable source of content validation for the JAAT-3. To achieve that, Cronbach's alpha for reliability (α) was calculated for all the clinically, linguistically, and culturally related items in Questionnaire 2 (validation of test content based on SMEs ratings of the JAAT-3) using Likert scales. Additionally, questionnaire items using a Likert scale that are presented in Table 9 were analyzed to compute an estimation of reliability. Cronbach's alpha reliability values higher than 0.7 are desirable,

with values of 0.8 and higher considered to indicate robust reliability (Pallant, 2013; Taber, 2018). Table 13 presents results that range from .78 to .89. These values are considered high and indicate strong levels of reliability for Questionnaire 2 (validation of test content based on SMEs ratings of the JAAT-3).

Table 13

Reliability of Questionnaire Items

Group of Questionnaire Items	α
Items appraising clinical aspects of the JAAT on the subtest level	0.899
Items appraising linguistic relevance of the JAAT on the subtest level	0.897
Items appraising cultural sensitivity of the JAAT on the subtest level	0.895
Items appraising overall SMEs impressions of the JAAT on the subtest level	0.880
Items appraising clinical aspects of the JAAT as a testing unit	0.780

CHAPTER V

DISCUSSION

The pressing need for formal tests to provide consistent clinical information has led to the development of an array of aphasia tests, especially for English-speaking populations. An aphasia test developer has several options when it comes to the type of aphasia test to develop but many have limited use and known shortcomings. While screening tools are easy and quick to administer (El Hachioui et al., 2017), their goal being to highlight the existence of aphasia without providing further information. (Blake, McKinney, Treece, Lee, & Lincoln, 2002). Functional aphasia instruments are presented with poor theoretical foundation, weakly defined variables, excessive dependence on relatives/caregivers of the PWA, lack of cultural sensitivity, and controversial reliability and validity values (Prins & Bastiaanse, 2004; Sacchett & Marshall, 1992). Aphasia tests for specific aspects of language and subgroups can be used with a set of tests because they cover a limited aspect of the disorder or severity level (Spren & Risser, 2003). Comprehensive aphasia tests offer much more information for clinical use by covering multiple aspects of communication that are necessary for diagnosis, prognosis, and treatment (Ivanova & Hallowell, 2013). Clinician-friendly tests have traits that are valued by clinicians that (a) take a relatively short time to administer, (b) can serve a wide spectrum scale of severity, and (c) can be suitable for use in all health care settings (Marshall & Wright, 2007).

The JAAT-3 represents a first step in the development of a comprehensive aphasia assessment battery. Although the JAAT's design was influenced by the BDAE-3 (Goodglass et al., 2001), it is a new test that takes into consideration being sensitive to cultural and linguistic aspects of the Jordanian population. It is these factors, so markedly different from other languages and cultures that make designing a new tool for PWA in Jordan the better option instead of adapting an existing instrument designed for a population with a significantly different language and culture (Hambleton & Patsula, 1999). This study followed structural development procedures for designing an aphasia test by focusing on the analysis of the test itself and its items through working with substantive experts in order to achieve content validity (Mislevy, 2007). The current study documents three steps that were followed to develop testing content: (a) verification of visual stimuli by non-brain-damaged (NBD) Jordanian Arabic speakers in JAAT-1; (b) identifying inadequate items in the JAAT-2 by administering the test on senior NBD Jordanian Arabic speakers; and (c) validating of test content based on SMEs' ratings of the JAAT-3. The construction process integrates content validation by carefully examining items, subtests, and parts of the JAAT. This framework of development and content validation can facilitate subsequent test development and validation stages in the future (Carretero-Dios, Pérez, & Buela-Casal, 2009).

It is important to carefully design test images for aphasia assessments to improve the validity of the assessment (Heuer & Hallowell, 2007). The first experiment in this study examined the clarity of pictures drawn for the JAAT-1 by presenting them to a group of NBD Jordanian-Arabic speakers. As predicted, most pictures in the JAAT-1 had

high clarity rates. These results indicate that pictures used in the preliminary version of the JAAT were easily recognizable by the vast majority of participants in Experiment One (Questionnaire 1: Verification of the JAAT-1 visual stimuli by NBD Jordanian Arabic speakers) and can be considered adequate visual representation of the words they target. The results of Experiment One (Questionnaire 1: Verification of the JAAT-1 visual stimuli by NBD Jordanian Arabic speakers) also indicates that the pictures posed no cultural challenges for the group as evidenced by the high values of agreement found across the participants.

It is erroneous to assume that individuals with no language impairments will obtain perfect scores on aphasia tests (Ross & Wertz, 2004) and Experiment Two (identifying inadequate items in the JAAT-2 by administering the test on senior NBD Jordanian Arabic speakers) was designed to identify particularly challenging items for a group of NBD Jordanian-Arabic speakers above the age of 60. The reasoning is that if an item is particularly challenging for individuals who have no communication impairments, then it will be even more challenging for a PWA. This may render an item useless if most PWA performed incorrectly on it. Although incorrect responses were reported for several items in the JAAT, subjects did perform correctly on most of them. *Expository speech* assessing subtests in the *Verbal Fluency* part, and items in the *Picture Naming* part proved particularly difficult for the group. These items required modification to better depict target words for the modified version of the JAAT and for further investigation in Experiment Three (Questionnaire 2: Validation of test content based on the SME rating of the JAAT-3).

A researcher's determination of item suitability is subjective and unreliable; therefore, acquiring statistical measures for an instrument is the objective option for content validation (Kyngas et al., 2000). Establishing content validity through expert ratings and inferences based on index value is a reliable method of examining an instrument's content validity (Hyrkas et al., 2003). Experiment Three (Questionnaire 2: Validation of test content based on SME rating of the JAAT-3) was set up to evaluate the JAAT's content based on three parameters: clinical applicability, linguistic relevance, and cultural sensitivity.

To be able to examine the clinical applicability of the JAAT, all questionnaire items examining this parameter were analyzed. This included 12 questionnaire items on the subtest level, four questionnaire items on the part level, and seven questionnaire items on the assessment as a whole test. On the subtest level, all but one subtest elicited positive responses with high agreement on content validity. The subtest assessing *expository speech* was the only subtest in the JAAT reporting lower CVI values on all parameters. This is one of the most important subtests in the JAAT due to the importance of evaluating spontaneous speech in a semi-structured manner by asking a PWA to describe a picture containing several actions or events. Spontaneous speech analysis is one of the most prominent characteristics of aphasia and considered a crucial ability to assess in any aphasia test (Prins & Bastiaanse, 2004). The fact that this subtest failed to achieve the criterion CVI value in Linguistic relevance and cultural sensitivity may explain the SME's less than favorable impression on its clinical applicability. The results indicate that the SMEs viewed the subtest's *Guest Room* picture to need more

modifications to enhance its linguistic and cultural suitability. The use of pictures is useful, so dropping the *Guest Room* picture is not be a viable option but enhancing the fidelity of the pictures and adding more cultural components will be required in the next stage for this subtest. All the questionnaire items targeting the four parts of the JAAT reported high positive responses from raters with high agreement of at least 85%. Raters viewed the JAAT to contain an adequate number of items to assess the four major domains of aphasia.

Raters exhibited high positive agreement regarding the usefulness of the JAAT's content however, this was not the case when considering the JAAT's usefulness in determining severity rates and identifying other concomitant disorders such as apraxia. Apraxia of speech is a neurogenic speech disorder that impairs the capacity to plan or program sensorimotor commands necessary for directing oro-facial movements for production of speech (Duffy, 2013). While Aphasia is more neurologically related to linguistic capacity and apraxia is more involved in speech motor programming, it is difficult to differentiate between the two disorders, especially if they coexist. Yet, it is important to state that aphasia tests are neither designed to diagnose apraxia, nor should they be. The same can also be said about dysarthria, a disorder that reduces the range strength of muscles involved in speech production. This should not compromise the content validity of the JAAT because the JAAT is designed to diagnose aphasia. That said, increasing the number of items in the Repetition part of the JAAT, as in the case of the BDAE-3 (Goodglass et al., 2001), may help clinicians in identifying some deficits in speech production and direct them to further specialized examinations of this area.

Determining severity levels in PWA is important for the allocation of clinical services. Unfortunately, the experts in this study did not provide further elaboration to explain where the JAAT's content was lacking in relation to severity levels. A test can lack the ability to categorize examinees according to severity of the disorder by virtue of being too easy or too hard. If the test is too easy, mild levels of aphasia may perform too well to be diagnosed and severe levels of aphasia may be categorized as moderate. If a mean performance score of 104.75 on a test that an individual can have a maximum score of 112 is an indication, then the JAAT can be assumed to be too easy for PWA. Additionally, 80% of subjects in Experiment Two had scores exceeding 100 on the JAAT. Adding more difficult items to the JAAT may contribute to a better capacity to differentiate severity levels in PWA. Of course, future standardization and norming research using subjects with no brain damage and with aphasia can provide empirical indications as to whether that capacity is achieved. Overall, and as predicted, raters judged most of the JAAT's content to be clinically applicable for Jordanian PWA with high average CVI of above 80%.

Highly important in our understanding of language is to draw an understanding of cultural differences that each language may impose on its users; this is more evident in our understanding of language impairments caused by aphasia. One of the most drastic instrument changes that is needed to overcome construct bias is removing specific items that are not suitable for a specific culture (van de Vijver & Hambleton, 1996). As hypothesized, raters reported that most subtests are linguistically relevant and culturally sensitive for the Jordanian population. This was supported by the 90% agreement rate

within raters that the JAAT as a whole is both linguistically and culturally sensitive. Apart from the *expository speech* assessing subtest, all subtests were rated as culturally sensitive to the Jordanian population with CVI rates above 80%. Another subtest that had a lower CVI rate in linguistic relevance was the subtest assessing *comprehension of sentences and paragraphs*. This subtest contains sentences and paragraphs that are designed to be read to an examinee followed by yes/no questions to respond to. Although the subtest scored well on clinical and cultural parameters, it had a CVI rate less than 80% and mean of responses below the overall mean. Further analysis and modification are required for this subtest. Adding more Jordanian-specific linguistic content, as well as items such as short stories influenced by folkloric themes, should be considered.

The results from Questionnaire 2 (validation of test content based on SMEs' ratings of the JAAT-3) reported in this study can be regarded as reliable. The subject matter experts were Jordanian SLPs who passed certification, with most having graduate degrees and reporting working with PWAs for several years. The number of experts ($n = 21$) that participated in the study also reinforces the reliability of the results, as Lynn (1986) recommends five to ten experts for test validation. The high Cronbach's alphas that were computed corroborate the overall reliability of the study.

Clinical Implications

Comprehensive aphasia assessments are an important part of diagnosis and treatment of PWA. The results of these assessments help determine the language deficits and play a vital role in rehabilitative service allocation. Jordanian health care providers, especially SLPs, are faced with challenges when gathering assessment information for

PWA. This is particularly true in the absence of formal tests that can provide useful clinical information while considering cultural and linguistic aspects of the population. The need for a formal comprehensive test is paramount; it represents an excellent assessment tool that is vital to collect and document a meaningful understanding of the communicative skills PWA have and assists professionals in research and providing appropriate rehabilitation (David, 1990). Although some tests have outdated theoretical bases such as emphasis on classifications derived from the localization theory or a unified language or the stimulation hypothesis model (Byng et al., 1990; Howard et al., 2010), they and the theories behind them still represent valuable clinical tools that are useful (Hillis, 2007). This is not to say that a formal comprehensive test provides an all-encompassing solution for evaluating PWA. Collecting information about a client's language abilities will always need to make use of careful behavioral observation and informal assessment items (Holland, 1982; Holland & Fridriksson, 2001; Marshall, 1997; Marshall & Wright, 2007) as well as gathering information from caregivers.

Although JAAT-3 is by no means a finished product and needs to go through more item and test development and standardization in the future, it does show promising potential in terms of clinical applicability, linguistic relevance, and cultural sensitivity. It contains items that are important in assessing the four major language components affected by aphasia: verbal fluency, auditory comprehension, repetition, and naming. With no formal testing instruments available in Jordan, the JAAT (version 3) will be a useful tool once it passes through its final stages of development and validation.

Limitations and Future Directions

There were some limitations associated with this study. The sample size in Experiment Two (identifying inadequate items in the JAAT-2 by administering the test on senior NBD Jordanian Arabic speakers) was relatively small. The items used were chosen subjectively due to lack of lexical databases for Jordanian Arabic at the time of test development. This may have caused some items to be inadequate for the Jordanian population. Efforts leading to the creation of lexicon databases for populations in the Middle East will help in the selection of appropriate linguistic stimuli. Although electronic questionnaires are an inexpensive means of gathering information from content experts, only a few raters gave elaborated responses on subtests they found less than adequate. In future individual interviews or expert panel meetings following a standard protocol of item analysis should provide appropriate feedback regarding the test content. Item analysis would not only confirm the validity of the test from a micro perspective, but also allow better comparisons across all parts of the test (verbal fluency, auditory comprehension, repetition, and naming) and contribute to the overall content validation.

Summary and Conclusions

The purpose of this study was to develop a framework to design and analyze an aphasia test specifically designed for the Jordanian Arabic-speaking population and improve the test's capacity to collect clinical information while maintaining linguistic and cultural sensitivity. This was carried out by three experiments aimed at validating visual stimuli, identifying inadequate items, and using SMEs to judge the test's content. As expected, the pictures used on the JAAT proved to be recognizable by Jordanian NBD

individuals varying in age, sex, and education. Most of the items on the JAAT generated correct responses from the senior NBD individuals who took it and less than 1% of the items on the JAAT were found to be too difficult. Items that proved to be unidentifiable in the first and second experiments were modified or deleted.

Jordanian content experts in the field of speech-language pathology were asked to judge the JAAT's content. The three aspects of appraisal were clinical applicability and linguistic and cultural relevance. Most subtests of the JAAT scored consistently high on raters' agreement and content validation indices on all appraisal parameters. They were found to be clinically applicable, linguistically relevant, and culturally sensitive for the target population. While the methodology of content validation using expert judgment was very useful in indicating that most subtests of the JAAT showed strong performance on three parameters, it was also crucial in identifying two subtests that performed below acceptable criterion and that require further future modifications: expository speech (in Verbal Fluency) comprehension of sentences and paragraphs (in Auditory Comprehension). The raters found that all parts of the JAAT contain enough items to assess the language abilities they were designed to assess.

As a unit, the JAAT was reported to be suitably timed, linguistically and culturally sensitive for its target population, showed a strong indication of being a useful tool for gathering diagnostic and therapeutic information. Additionally, the JAAT was rated highly when it came to possible future preference for clinicians if it was made available to them. In contrast, the ability of the JAAT to identify other disorders such as dysarthria and apraxia, as well as determine severity levels of aphasia, were found to be

lacking. These findings represent improvement opportunities and reveal the benefit of such methodology in identifying strengths and weaknesses in test content. The consistency of responses on the test's rating tool, with a Cronbach's alpha of more than 0.8, indicated a high reliability value overall.

After following a framework of test design aimed at validating content used in the test, the results of this study show that the JAAT exhibits promising clinical, linguistic, and cultural attributes. Additionally, the results corroborate the merit of this content validation framework in test development. Continued work on the JAAT's content can be followed to in the shape of research and development to finalize this comprehensive assessment for use with Jordanian Arabic-speaking PWA.

REFERENCES

- Abo Ras, Y., Aref, S., El-Raghy, A., Gaber, O., & El-Maghraby, R. (2009). *Comprehensive Arabic Language test as tool for assessing delayed language impaired Egyptian children*. Retrieved from <http://www.alexorl.edu.eg/alexorlfiles/pptorl2009/014003.pdf>
- About 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.
- Achaston, S. K. (2010). [Test review of Western Aphasia Battery-Revised]. In R. A. Spies, J.F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements Yearbook*. Retrieved from <http://marketplace.unl.edu/buros/>
- Al-Thalaya, Z., Nilipour, R., Sadat Ghoreyshi, Z., Pourshahbaz, A., Nassar, Z., & Younes, M. (2018). Reliability and validity of bedside version of Arabic Diagnostic Aphasia Battery (A-DAB-1) for Lebanese individuals. *Aphasiology*, 32(3), 323–339.
- Al Yaari, S. A. S., & Almaflehi, N. (2013). Validity and Reliability of Assessment of Language-Related Functional Activities (ALFA): Evidence from Arab Aphasics. *Journal for the Study of English Linguistics*, 1(2), 121–141.
- Alyahya, R. S., & Druks, J. (2016). The adaptation of the object and action naming battery into Saudi Arabic. *Aphasiology*, 30(4), 463–482.

American Psychological Association, American Educational Research Association, National Council on Measurement in Education, & American Educational Research Association. Committee on Test Standards. (1966). *Standards for educational and psychological tests and manuals*. American Psychological Association.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. American Educational Research Association.

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Speech and Hearing Association. (2018). What is Aphasia. Retrieved March 15, 2018 from <http://www.asha.org/public/speech/disorders/Aphasia/>

American Speech and Hearing Association. (2019a). *Assessment tools, techniques, and data sources*. Retrieved from <https://www.asha.org/practice-portal/clinical-topics/late-language-emergence/assessment-tools-techniques-and- data-sources/>

American Speech and Hearing Association. (2019b). *Clinical topics: Aphasia*. Retrieved from <https://www.asha.org/Practice-Portal/Clinical-Topics/aphasia/>

- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1–15.
- Baines, K. A., Heeringa, H. M., & Martin, A. W. (1999). *Assessment of Language-related Functional Activities: ALFA*. Pro-Ed.
- Ball, M. J., Perkins, M. R., Müller, N., & Howard, S. (Eds.). (2008). *The handbook of clinical linguistics*. Cambridge: Blackwell.
- Bastiaanse, R., Maas, E., & Rispens, J. (2000). *Werkwoorden- en Zinnetest (WEZT)*. Lisse, The Netherlands: Swets & Zeitlinger.
- Bikel, D. M. (2002, March). Design of a multi-lingual, parallel-processing statistical parsing engine. In *Proceedings of the second international conference on Human Language Technology Research* (pp. 178-182).
- Blake, H., McKinney, M., Treece, K., Lee, E., & Lincoln, N. B. (2002). An evaluation of screening measures for cognitive impairment after stroke. *Age and ageing*, 31(6), 451-456.
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016, January 1). Speech and language therapy for aphasia following stroke. *The Cochrane Database of Systematic Reviews*, 2016(6).
doi:10.1002/14651858.CD000425.pub4
- Brandenburg, C., Worrall, L., Copland, D., Power, E., & Rodriguez, A. D. (2016). The development and accuracy testing of CommFit™, an iPhone application for individuals with aphasia. *Aphasiology*, 30(2-3), 320–338.
doi:10.1080/02687038.2015.1028329

- Bridges, S. J. (2004). Multicultural issues in augmentative and alternative communication and language: Research to practice. *Topics in Language Disorders, 24*(1), 62–75. doi:10.1097/00011363-200401000-00007
- Bruce, C., & Edmundson, A. (2010, January 1). Letting the CAT out of the bag: A review of the Comprehensive Aphasia Test. Commentary on Howard, Swinburn, and Porter, “Putting the CAT out: What the Comprehensive Aphasia Test has to offer.” *Aphasiology, 24*(1), 79–93. doi:10.1080/02687030802453335
- Brust, J. C., Shafer, S. Q., Richter, R. W., & Bruun, B. (1976). Aphasia in acute stroke. *Stroke, 7*(2), 167–174. <https://www.ncbi.nlm.nih.gov/pubmed/1265809>
- Byng, S., Kay, J., Edmundson, A., & Scott, C. (1990, January 1). Aphasia tests reconsidered. *Aphasiology, 4*(1), 67–91. doi:10.1080/02687039008249055
- Carretero-Dios, H., & Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology, 7*(3), 863–882. Retrieved from http://www.aepc.es/ijchp/carretero_es.pdf
- Carretero-Dios, H., Pérez, C., & Buela-Casal, G. (2009). Content validity and metric properties of a pool of items developed to assess humor appreciation. *The Spanish Journal of Psychology, 12*(2), 773–787. doi:10.1017/S1138741600002146
- Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex, 44*(8), 953–961. doi:10.1016/j.cortex.2008.04.002

- Chiang, D., Diab, M., Habash, N., Rambow, O., & Shareef, S. (2006). Parsing Arabic dialects. In *11th Conference of the European Chapter of the Association for Computational Linguistics*. Retrieved from http://www.cs.columbia.edu/nlp/papers/2006/chiang_al_06.pdf
- Coppens, P. (2016). *Aphasia and related neurogenic communication disorders*. Burlington, MA: Jones & Bartlett.
- Croquelois, A., & Bogousslavsky, J. (2011). Stroke aphasia: 1,500 consecutive cases. *Cerebrovascular Diseases, 31*(4), 392–399. doi:10.1159/000323217
- Cunningham, R., Farrow, V., Davies, C., & Lincoln, N. (1995). Reliability of the assessment of communicative effectiveness in severe aphasia. *International Journal of Language & Communication Disorders, 30*(1), 1–16. doi:10.3109/13682829509031319
- Cureton, E. E. (1951). Validity. In E. F. Lindquist (Ed.), *Educational measurement*. (pp. ??-??). Washington, DC: American Council on Education.
- David, R. (1990). Aphasia assessment: The acid tests. *Aphasiology, 4*(1), 150–154. doi:10.1002/9780470699171.ch4c
- Davis, G. A. (2007). *Aphasiology: Disorders and clinical practice*. Boston, MA: Pearson.
- Davis, A. S., & Finch, H. (2010). [Test review of Western Aphasia Battery-Revised]. In R. A. Spies, J.F. Carlson, & K. F. Geisinger (Eds.), *The eighteenth mental measurements Yearbook*. Retrieved from <http://marketplace.unl.edu/buroos/>

- del Rosario Basterra, M., Trumbull, E., & Solano-Flores, G. (Eds.). (2011). *Cultural validity in assessment: Addressing linguistic and cultural diversity*. New York, NY: Routledge.
- Delgado-Rico, E., Carrctero-Dios, H., & Ruch, W. (2012). Content validity evidences in test development: An applied perspective. *International Journal of Clinical and Health Psychology, 12*(3), 449–460.
- DeMaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 89–108). Hoboken, NJ: John Wiley & Sons.
- Dickey, L., Kagan, A., Lindsay, M. P., Fang, J., Rowland, A., & Black, S. (2010, February 1). Incidence and profile of inpatient stroke-induced aphasia in Ontario, Canada. *Archives of Physical Medicine and Rehabilitation, 91*(2), 196–202.
doi:10.1016/j.apmr.2009.09.020
- Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management*. Elsevier Health Sciences.
- Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. American Guidance Service.
- Dronkers, N. F., & Baldo, J. V. (2010). Language: Aphasia. In *Encyclopedia of neuroscience*. Elsevier Ltd.
- Druks, J., & Masterson, J. (2000). *An Object and Action Naming Battery*. London: Psychology Press.

- Ebel, R. (1961). Must all tests be valid? *American Psychologist*, *16*(10), 640–647.
doi:10.1037/h0045478
- El Hachoui, H., Visch-Brink, E. G., de Lau, L. M., van de Sandt-Koenderman, M. W.,
Nouwens, F., Koudstaal, P. J., & Dippel, D. W. (2017). Screening tests for
aphasia in patients with stroke: A systematic review. *Journal of
Neurology*, *264*(2), 211–220. doi:10.1007/s00415-016-8170-8
- El-Hajj, M., Salameh, P., Rachidi, S., & Hosseini, H. (2016). The epidemiology of stroke
in the Middle East. *European Stroke Journal*, *1*(3), 180–198.
doi:10.1177/2396987316654338
- Ellis, C., Dismuke, C., & Edwards, K. K. (2010). Longitudinal trends in aphasia in the
United States. *NeuroRehabilitation*, *27*(4), 327–333.
doi:10.3233/NRE-2010-0616
- Ellis, C., Hardy, R. Y., Lindrooth, R. C., & Peach, R. K. (2017). Rate of aphasia among
stroke patients discharged from hospitals in the United States. *Aphasiology*, *32*(9),
1075–1086. doi:10.1080/02687038.2017.1385052
- Enderby, P., Wood, V., & Wade, D. (1987). *Frenchay Aphasia Screening Test (FAST):
Test Manual*. Whurr Publishers.
- Enderby, P. M., Wood, V. A., Wade, D. T., & Hewer, R. L. (1986). The Frenchay
Aphasia Screening Test: a short, simple test for aphasia appropriate for non-
specialists. *International Rehabilitation Medicine*, *8*(4), 166–170.
doi:10.3109/03790798709166209

- Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., . . . Lyrer, P. A. (2006, January 1). 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
- Feigin, V. L., Mensah, G. A., Norrving, B., Murray, C. J., & Roth, G. A. (2015). Atlas of the global burden of stroke (1990-2013): The GBD 2013 study. *Neuroepidemiology*, *45*(3), 230–236. doi:10.1159/000441106
- Flowers, H. L., Skoretz, S. A., Silver, F. L., Rochon, E., Fang, J., Flamand-Roze, C., & Martino, R. (2016, December 1). Poststroke aphasia frequency, recovery, and outcomes: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation*, *97*(12), 2188–2201. doi:10.1016/j.apmr.2016.03.006
- Frattali, C. M., Thompson, C. M., Holland, A. L., Wohl, C. B., & Ferketic, M. M. (1995). The FACS of life ASHA facs—A functional outcome measure for adults. *ASHA*, *37*(4), 40–46. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/7733997>
- Fridriksson, J., den, O. D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., . . . Bonilha, L. (2018, January 1). Anatomy of aphasia revisited. *Brain*, *141*(3), 848–862.
- Fyndanis, V., Lind, M., Varlokosta, S., Kambanaros, M., Soroli, E., Ceder, K., . . . Gavarró, A. (2017). Cross-linguistic adaptations of the Comprehensive Aphasia Test: Challenges and solutions. *Clinical Linguistics & Phonetics*, *31*(7-9), 697–710. doi:10.1080/02699206.2017.1310299

- Gales, L. M. (2003). Linguistic sensitivity in cross-cultural organisational research: Positivist/post-positivist and grounded theory approaches. *Language and Intercultural Communication*, 3(2), 131–140.
- Garrett, H. E. (1937). *Statistics in psychology and education*. Oxford, England: Longmans, Green.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304–312. doi:10.1037/1040-3590.6.4.304
- Gialanella, B. (2011, February 1). Aphasia assessment and functional outcome prediction in patients with aphasia after stroke. *Journal of Neurology*, 258(2), 343–349. doi:10.1007/s00415-010-5868-x
- Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination Record Booklet*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Goodglass, H., Kaplan, E., & Baressi, B. (2001). *Boston Diagnostic Aphasia Examination* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.
- Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders*. Philadelphia, PA: Lea & Febiger.
- Goodglass, H., & Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*. Lea & Febiger.
- Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20(3), 269–274. doi:10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

- Gregory, R. J. (2007). *Psychological testing: History, principles, and applications*. Boston, MA: Pearson/Allyn and Bacon.
- Haley, K. L., Roth, H., Grindstaff, E., & Jacks, A. (2011). Computer-mediated assessment of intelligibility in aphasia and apraxia of speech. *Aphasiology*, 25(12), 1600–1620. doi:10.1080/02687038.2011.628379
- Hallowell, B., & Chapey, R. (2008). Introduction to language intervention strategies in adult aphasia. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (5th ed., pp. 3–19). Baltimore, MD: Lippincott Williams & Wilkins.
- Hambleton, R. K. (1996). *Guidelines for adapting educational and psychological tests*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, NY. Retrieved from <https://files.eric.ed.gov/fulltext/ED399291.pdf>
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (Eds.). (2004). *Adapting educational and psychological tests for cross-cultural assessment*. Psychology Press.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45(1-3), 153–171. doi:10.1023/A:1006941729637
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of*

- Applied Testing Technology*, 1(1), 1–13. Retrieved from
<http://www.jattjournal.com/index.php/atp/article/view/48345/39215>
- 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 of Neurology, Psychiatry, and Neurosurgery*, 39, 381–395.
- Hegde, M. N., & Freed, D. B. (2011). *Assessment of communication disorders in adults*. Plural Publishing.
- Helm-Estabrooks, N., Ramsberger, G., Morgan, A. R., & Nicholas, M. (1989). *BASA: Boston Assessment of Severe Aphasia*. Riverside Publishing Company.
- Hersh, D., Wood, P., & Armstrong, E. (2018). Informal aphasia assessment, interaction and the development of the therapeutic relationship in the early period after stroke. *Aphasiology*, 32(8), 876–901. doi:10.1080/02687038.2017.1381878
- Heuer, S., & Hallowell, B. (2007). An evaluation of multiple-choice test images for comprehension assessment in aphasia. *Aphasiology*, 21(9), 883–900. doi:10.1080/02687030600695194
- Heuer, S., Ivanova, M. V., & Hallowell, B. (2017, January 1). More than the verbal stimulus matters: Visual attention in language assessment for people with aphasia using multiple-choice image displays. *Journal of Speech, Language, and Hearing Research*, 60(5), 1348–1361. doi:10.1044/2017_JSLHR-L-16-0087

- Hillis, A. E. (2007, July 10). Aphasia: Progress in the last quarter of a century. *Neurology*, *69*(2), 200–213. doi:10.1212/01.wnl.0000265600.69385.6f
- Holland, A. L. (1982). Observing functional communication of aphasic adults. *Journal of Speech and Hearing Disorders*, *47*(1), 50–56. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/7176575>
- Holland, A., & Fridriksson, J. (2001). Aphasia management during the early phases of recovery following stroke. *American Journal of Speech-Language Pathology*, *10*(1), 19–28. doi:10.1044/1058-0360(2001/004)
- Howard, D., Swinburn, K., & Porter, G. (2010, January 1). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. *Aphasiology*, *24*(1), 56–74. doi:10.1080/02687030802453202
- Huber, W., Grande, M., & Springer, L. (2005). *Aachener Sprachanalyse (ASPA)*. Aachen: Delta Systems.
- Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). *Der aachener aphasia test*. Gottingen: Verlag fur Psychologie Hogrefe.
- Hughes, L. C. (1998). Development of an instrument to measure caring peer group interactions. *Journal of Nursing Education*, *37*(5), 202-207
- Hussmann, K., Grande, M., Meffert, E., Christoph, S., Piefke, M., Willmes, K., & Huber, W. (2012). Computer-assisted analysis of spontaneous speech: Quantification of basic parameters in aphasic and unimpaired language. *Clinical Linguistics & Phonetics*, *26*(8), 661–680. doi:10.3109/02699206.2012.683672

- Hyrkäs, K., Appelqvist-Schmidlechner, K., & Oksa, L. (2003). Validating an instrument for clinical supervision using an expert panel. *International Journal of Nursing Studies*, *40*(6), 619–625. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/12834927>
- Inatomi, Y., Yonehara, T., Omiya, S., Hashimoto, Y., Hirano, T., & Uchino, M. (2008). Aphasia during the acute phase in ischemic stroke. *Cerebrovascular Diseases*, *25*(4), 316–323. doi:10.1159/000118376
- Irwing, P., & Hughes, D. J. (2018). Test development. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 1–47). Wiley-Blackwell.
- Ivanova, M., & Hallowell, B. (2009, January 1). Short form of the Bilingual Aphasia Test in Russian: Psychometric data of persons with aphasia. *Aphasiology*, *23*(5), 544–556. doi:10.1080/02687030701800784
- Ivanova, M. V., & Hallowell, B. (2013, January 1). A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. *Aphasiology*, *27*(8), 891–920. doi:10.1080/02687038.2013.805728
- Johnson, M. L., Taub, E., Harper, L. H., Wade, J. T., Bowman, M. H., Bishop-McKay, S., . . . Uswatte, G. (2014). An enhanced protocol for constraint-induced aphasia therapy II: A case series. *American Journal of Speech-Language Pathology*, *23*(1), 60–72. doi:10.1044/1058-0360(2013/12-0168)

- Kadojić, D., Rostohar Bijelić, B., Radanović, R., Porobić, M., Rimac, J., & Dikanović, M. (2012). Aphasia in patients with ischemic stroke. *Acta Clinica Croatica*, *51*(2), 221–224. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23115946>
- Kauhanen, M. L., Korpelainen, J. T., Hiltunen, P., Määttä, R., Mononen, H., Brusin, E., . . . Myllylä, V. V. (2000). Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovascular Diseases*, *10*(6), 455–461. doi:10.1159/000016107
- Kempler, D. (2005). *Neurocognitive disorders in aging*. Thousand Oaks, CA: Sage.
- Kertesz, A. (2006). Western aphasia battery-revised (WAB-R): Examiner's manual. *PsychCorp. Harcourt Assessment Incorporation*.
- Khoja, M. A. (2017, November 24). A survey of formal and informal assessment procedures used by speech-language pathologists in Saudi Arabia. *Speech, Language and Hearing*, *6*, 1–9. doi:10.1080/2050571X.2017.1407620
- Khwaileh, T., Body, R., & Herbert, R. (2014, December 1). A normative database and determinants of lexical retrieval for 186 Arabic nouns: Effects of psycholinguistic and morpho-syntactic variables on naming latency. *Journal of Psycholinguistic Research*, *43*(6), 749–769. doi:10.1007/s10936-013-9277-z
- Khwaileh, T. A., Mustafawi, E., Howard, D., & Herbert, R. (2016). An aphasia battery for Qatari/Gulf Arabic. *Frontiers in psychology*.
- Kissela, B. M., Khoury, J. C., Alwell, K., Moomaw, C. J., Woo, D., Adeoye, O., . . . Broderick, J. P. (2012). Age at stroke: temporal trends in stroke incidence in a large, biracial population. *Neurology*, *79*(17), 1781–1787.

- Krishnamurthi, R. V., Moran, A. E., Feigin, V. L., Barker-Collo, S., Norrving, B., Mensah, G. A., . . . Johnson, C. O. (2015). Stroke prevalence, mortality and disability-adjusted life years in adults aged 20-64 years in 1990-2013: Data from the global burden of disease 2013 study. *Neuroepidemiology*, *45*(3), 190–202. doi:10.1159/000441098
- Kyngäs, H. A., Skaar-Chandler, C. A., & Duffy, M. E. (2000). The development of an instrument to measure the compliance of adolescents with a chronic disease. *Journal of Advanced Nursing*, *32*(6), 1499–1506. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11136419>
- Kyrozis, A., Potagas, C., Ghika, A., Tsimpouris, P. K., Virvidaki, E. S., & Vemmos, K. N. (2009). Incidence and predictors of post-stroke aphasia: The Arcadia Stroke Registry. *European Journal of Neurology*, *16*(6), 733–739. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/19475755>
- Lazar, R. M., Speizer, A. E., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2008). Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery and Psychiatry*, *79*(5), 530–534. doi:10.1136/jnnp.2007.122457
- Leung, S. F., & Arthur, D. (2000). The alcohol use disorders identification test (AUDIT): validation of an instrument for enhancing nursing practice in Hong Kong. *International Journal of Nursing Studies*, *37*(1), 57–64. doi:10.1016/S0020-7489(99)00052-8

- Lin, L. C., Chen, M. Y., Chen, Y. C., & Portwood, M. J. (2001). Psychometrics of a Chinese translation of the swallowing questionnaire. *Journal of Advanced Nursing*, 34(3), 296–303. <https://www.ncbi.nlm.nih.gov/pubmed/11328434>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/3640358>
- Marczinski, C. A., & Kertesz, A. (2006, June 1). Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer’s disease. *Brain and Language*, 97(3), 258–265. doi:10.1016/j.bandl.2005.11.001
- Marshall, R. C. (1997). Aphasia treatment in the early postonset period: Managing our resources effectively. *American Journal of Speech-Language Pathology*, 6(1), 5–11. doi:10.1044/1058-0360.0601.05
- Marshall, R. C., & Wright, H. H. (2007, November 1). Developing a clinician-friendly aphasia test. *American Journal of Speech-Language Pathology*, 16(4), 295–315. doi:10.1044/1058-0360(2007/035)
- Masterson, J., & Druks, J. (1998). Description of a set of 164 nouns and 102 verbs matched for printed word frequency, familiarity and age-of-acquisition. *Journal of Neurolinguistics*, 11(4), 331–354. doi:10.1016/S0911-6044(98)00023-2
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.
- Mishra, A., Catchpole, K., & McCulloch, P. (2009). The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the

operating theatre. *BMJ Quality & Safety*, 18(2), 104–108.

doi:10.1136/qshc.2007.024760

Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, 36(8), 463–469.

doi:10.3102/0013189X07311660

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).

Papathanasiou, I., Coppens, P., & Potagas, C. (2013). *Aphasia and related neurogenic communication disorders*. Burlington, MA: Jones & Bartlett Learning.

Paradis, M., & Abidi, R. (1987). *Fahs hallat fuqdan taqat annotk aw el kalam ande mudjeedi loghatein aw akthar* (Maghrebian Arabic version). Hillsdale, NJ: Lawrence Erlbaum.

Paradis, M. (2011). Principles underlying the Bilingual Aphasia Test (BAT) and its uses. *Clinical Linguistics & Phonetics*, 25(6-7), 427–443.

doi:10.3109/02699206.2011.560326

Paradis, M., & El Halees, Y. M. (1989). *Bilingual aphasia test* (Jordanian Arabic version). Hillsdale, NJ: Lawrence Erlbaum.

Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Hillsdale, NJ: Lawrence Erlbaum.

Pedersen, P. M., Stig Jørgensen, H., Nakayama, H., Raaschou, H. O., & Olsen, T. S. (1995). Aphasia in acute stroke: Incidence, determinants, and recovery. *Annals of Neurology*, 38(4), 659–666. doi:10.1002/ana.410380416

Penny, A. M., Waschbusch, D. A., Klein, R. M., Corkum, P., & Eskes, G. (2009).

Developing a measure of sluggish cognitive tempo for children: Content validity,

factor structure, and reliability. *Psychological Assessment*, 21(3), 380–389.

doi:10.1037/a0016600

Plowman, E., Hentz, B., & Ellis, C. J. (2012, January 1). Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice*, 18(3), 689–694. doi:10.1111/j.1365-2753.2011.01650.x

Clinical Practice, 18(3), 689–694. doi:10.1111/j.1365-2753.2011.01650.x

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? *Sociological Methodology*, 24, 73–104. doi:10.2307/270979

Prins, R., & Bastiaanse, R. (2004). Analyzing the spontaneous speech of aphasic

speakers. *Aphasiology*, 18(12), 1075–1091. doi:10.1080/02687030444000534

Ross, K., & Wertz, R. (2004, January 1). Accuracy of formal tests for diagnosing mild aphasia: An application of evidence-based medicine. *Aphasiology*, 18(4), 337–

355. doi:10.1080/02687030444000002

Rutten, G.-J. (2017). *The Broca-Wernicke doctrine: A historical and clinical perspective on localization of language functions*. Springer International Publishing.

Ryglewicz, D., Hier, D. B., Wiszniewska, M., Cichy, S., Lechowicz, W., &

Czlonkowska, A. (2000). Ischemic strokes are more severe in Poland than in the

United States. *Neurology*, 54(2), 513–515. doi:10.1212/WNL.54.2.513

Sacchett, C., & Marshall, J. (1992). Functional assessment of communication:

Implications for the rehabilitation of aphasic people: Reply to Carol Frattali.

Aphasiology, 6(1), 95–100. doi:10.1080/02687039208248579

- Saidat, A. M. (2018). Jordanian Arabic: A study of the motivations for the intentionality in dialect change. *Theory and Practice in Language Studies*, 8(6), 580–587.
doi:10.17507/tpls.0806.04
- Sarno, M. T. (1997, July 1). Quality of life in aphasia in the first post-stroke year. *Aphasiology*, 11(7), 665–679. doi:10.1080/02687039708249414
- Sbordone, R. J., Saul, R. E., & Purisch, A. D. (2007). *Neuropsychology for psychologists, health care professionals, and attorneys* (3rd ed.). Boca Raton, FL: CRC Press.
doi:10.1201/9781420007138
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education*, 42(2), 123–149. Retrieved from <https://files.eric.ed.gov/fulltext/EJ868626.pdf>
- Seo, J., MacEntee, M., & Brondani, M. (2015). The use of subject matter experts in validating an oral health-related quality of life measure in Korean. *Health and Quality of Life Outcomes*, 13(1), 138. doi:10.1186/s12955-015-0335-0
- Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 19–37). Charlotte, NC: IAP Information Age Publishing.

- Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice*, 25(3), 27–34. doi:10.1111/j.1745-3992.2006.00065.x
- Slocumb, E. M., & Cole, F. L. (1991). A practical approach to content validation. *Applied Nursing Research*, 4(4), 192–195. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/1772253>
- SLTA Committee. (1977). *Standard language test of aphasia: Manual of directions* (2nd ed.). Tokyo: Homeido.
- Spreen, O., & Risser, A. H. (2003). *Assessment of aphasia*. Oxford: Oxford University Press.
- Stipancic, K. L., Borders, J. C., Brates, D., & Thibeault, S. L. (2019). Prospective investigation of incidence and co-occurrence of dysphagia, dysarthria, and aphasia following ischemic stroke. *American Journal of Speech-Language Pathology*, 28(1), 188–194. doi:10.1044/2018_AJSLP-18-0136
- Swinburn, K., Porter, G., & Howard, D. (2005). *The Comprehensive Aphasia Test*. Hove, UK: Psychology Press.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. doi:10.1007/s11165-016-9602-2
- Tojib, D. R., & Sugianto, L. F. (2006). Content validity of instruments in IS research. *Journal of Information Technology Theory and Application (JITTA)*, 8(3), 5.

- Tran, J., Mirzaei, M., Anderson, L., & Leeder, S. R. (2010). The epidemiology of stroke in the Middle East and North Africa. *Journal of the Neurological Sciences*, 295(1), 38–40. doi:10.1016/j.jns.2010.05.016
- Tsapkini, K., Vlahou, C. H., & Potagas, C. (2010). Adaptation and validation of standardized aphasia tests in different languages: Lessons from the Boston Diagnostic Aphasia Examination–Short Form in Greek. *Behavioural Neurology*, 22(3-4), 111–119. doi:10.3233/ben-2009-0256
- van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1(2), 89–99. doi:10.1027/1016-9040.1.2.89
- Wade, D. T., Hewer, R. L., David, R. M., & Enderby, P. M. (1986). Aphasia after stroke: natural history and associated deficits. *Journal of Neurology, Neurosurgery & Psychiatry*, 49(1), 11–16. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1028640/>
- West, J. F., & Sands, E. S., & Ross-Swain, D. (1998). *Bedside Evaluation and Screening Test, 2nd ed. (BEST-2)*. Austin, TX: Pro-Ed
- Willmes, K., & Poeck, K. (1993). To what extent can aphasic syndromes be localized? *Brain*, 116(Pt. 6), 1527–1540. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/8293285>
- Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018, January 1). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLOS ONE*, 13(6), e0199469. doi:10.1371/journal.pone.0192773

APPENDIX A

THE JORDANIAN-ARABIC APHASIA TEST (JAAT)

Part 1
Verbal Fluency Part

Subtest 1 of 3: Descriptive speech subtest

Instruction: conduct an informal exchange with the patient incorporating the following dialogue to elicit open-ended verbal responses from examinee. Record verbatim and tape record the dialogue.

Use scores 2 for correct response, 1 for incomplete, and 0 for no response.

Score	Examinee response	Items
		1. أيش أسمك؟ (what is your name?)
		2. كيف حالك؟ (how are you?)
		3. كيف الجو اليوم؟ (how is the weather today?)
		4. أيش صار معك؟ (what happened to you?)
		5. وين ساكن؟ (where do you live?)
		6. وين إحنا الآن؟ (where are we now?)
		7. أكلت اليوم؟ أيش أكلت؟ (have you eaten today? what did you eat?)
		7. كم لك في المستشفى؟ (how long have you been in hospital?)
		9. مين جابك هون عندي؟ (who brought you here to me?)
		10. ايش (كنت) بتشتغل؟ (what do/did you do?)

Score: _____

Subtest 2 of 3: Picture description subtest

Instructions: present picture and ask the examinee to describe it by telling you about everything he/she sees in the picture encouraging the examinee to talk about all the aspects. Point to neglected areas making sure that they are within his/her visual field. Direct the examinee's attention to aspects within the picture he had not talked about to encourage a more complete response. Tape and record the sample. Write down observations. Record examinee's responses verbatim

1. Total number of utterances: _____
 2. Empty utterances: _____
 3. Subclausal utterances: _____
 4. Single clause utterances: _____
 5. Multi-clause utterances: _____
- Complexity index (clauses per utterance): _____



Score: _____

Subtest 3 of 3: Picture sequence description subtest

Instruction: Present the two sequenced pictures and ask the examinee to describe the events occurring in the pictures, probe for more details. Follow the same instructions for scoring for previous picture description subtest.

1. Total number of utterances: _____
 2. Empty utterances: _____
 3. Subclausal utterances: _____
 4. Single clause utterances: _____
 5. Multi-clause utterances: _____
- Complexity index (clauses per utterance): _____



Score: _____

Part 2 of 4
Auditory Comprehension Part

Subtest 1 of 4: Comprehension of commands

Instruction: Read the below commands to the examinee. Click for each response and give it a score of one. Put the sum of scores in the last box

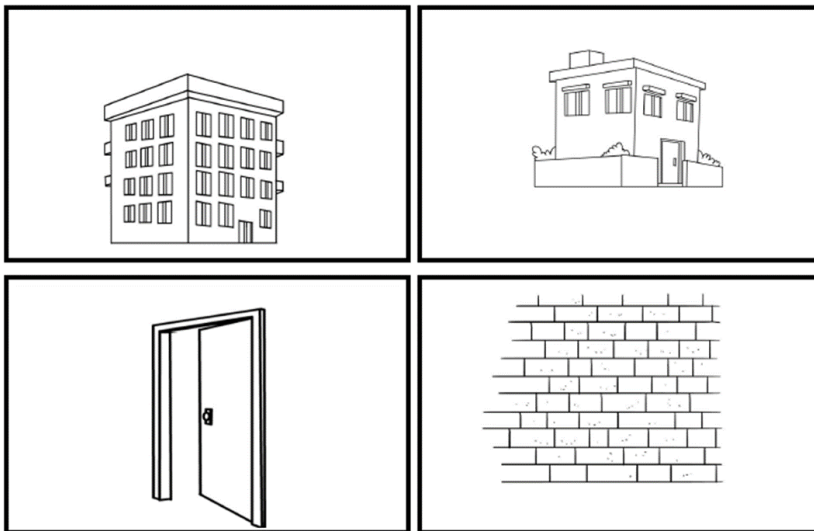
Sum of scores out of 9	Click for each sentence		Items in Arabic	Item Number
			غمض عينك Close your eyes	1
			خذ نفس Take a breath	2
			كح cough	3
			أمسك الكتاب/الدفتنر و من ثم إفتحه Hold the book and open it	4
			إمسك الساعة وقولي كم الوقت؟ Hold the watch and tell me the time	5
			إمسك القلم و أكتب اسمك بالهواء Hold the pen and write your name in the air	6
			أمسك المحرمة, أرفعها فوق, ثم شاور على الساعة Hold the tissue, lift it up and point at the watch	7
			امسك الورقة, إمزعها, ثم أعطيني القلم Hold the paper, tear it, and give me the pen	8
/9			أمسك القلم, ارسم خط على الورقة, ثم حط الساعة فوق الورقة. Hold the pen, draw a line with it on the paper, and put the watch on the paper	9

Subtest 2 of 4: Word comprehension

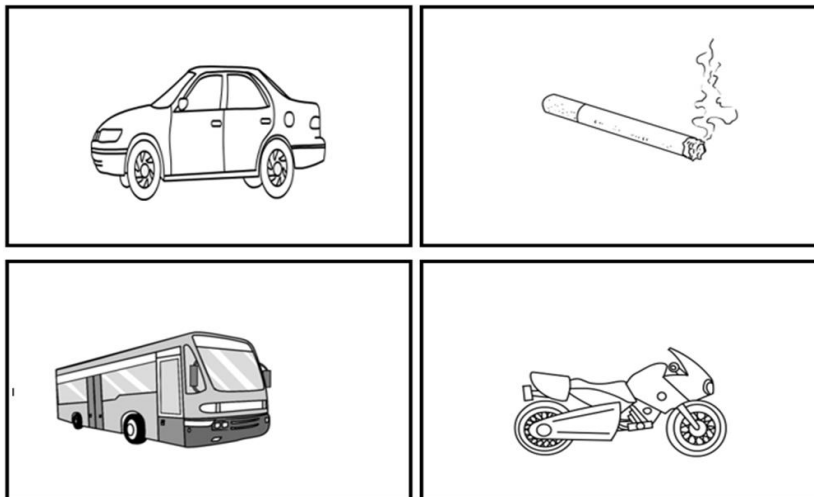
Instructions: Present pictures in a closed set of four and say: “show me the . . .” Instruct examinee to look directly at you when presenting the verbal stimuli. (Rationale: the words have been selected taking into consideration number of syllables, word frequency, semantic category, and cultural appropriateness).

One point is given for each correct response produced in less than five seconds and half a point for correct response produced after five seconds.

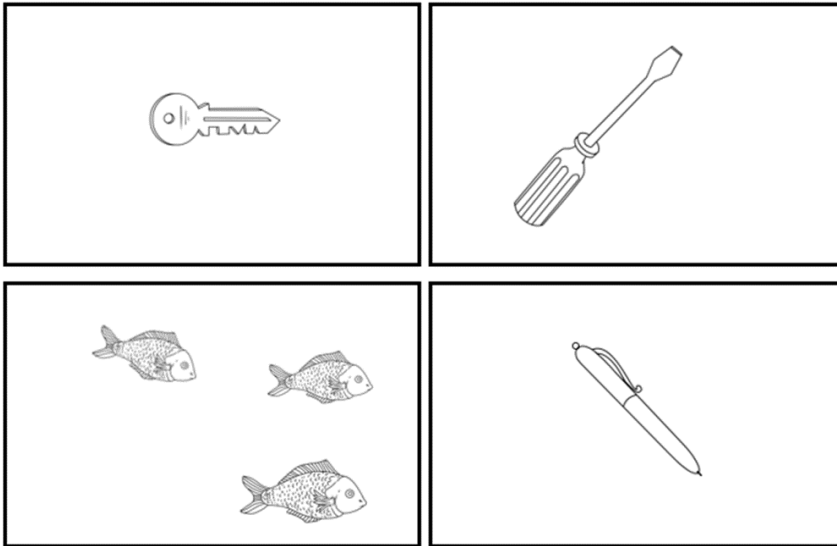
1 of 20 target words: بيت (House)



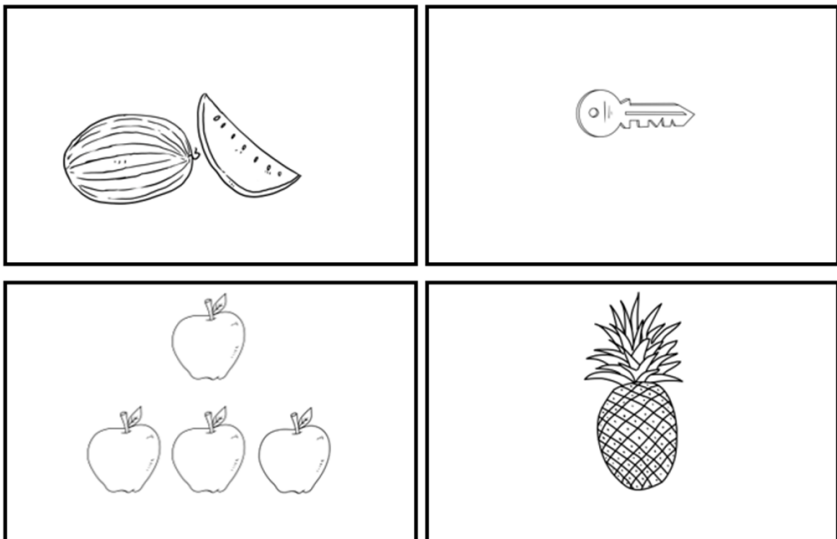
2 of 20 target words: يارة (Car)



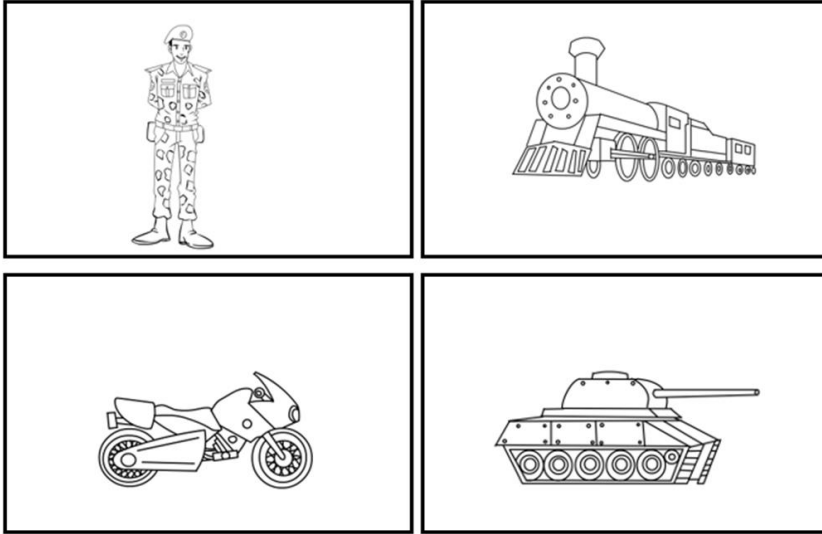
3 of 20 target words: مفك (Screwdriver)



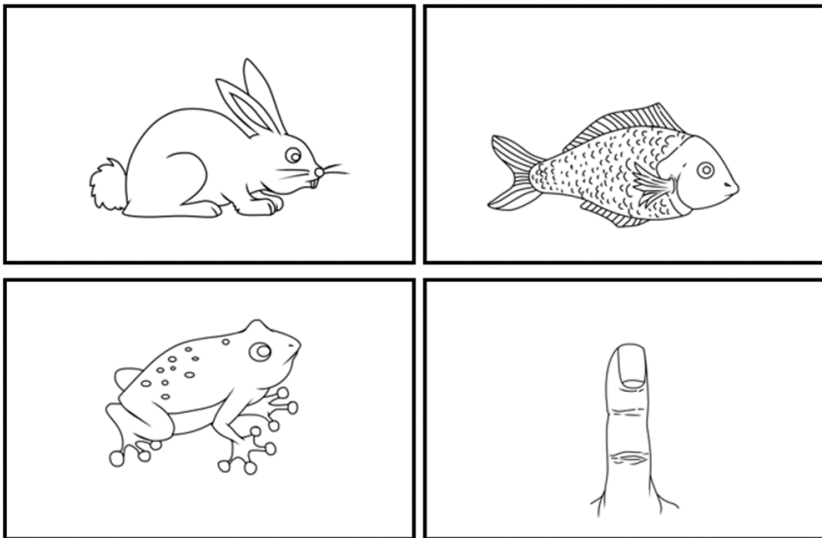
4 of 20 target words: تفاح (Apples)



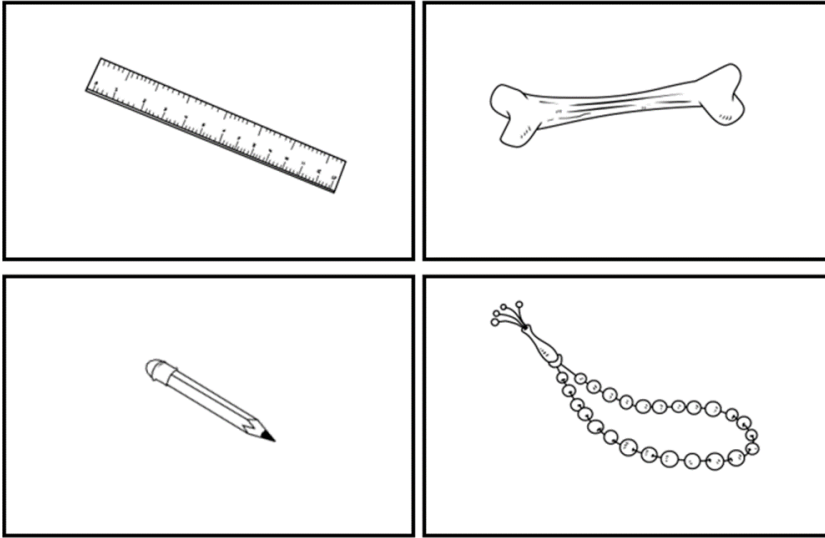
5 of 20 target words: دبابة (Tank)



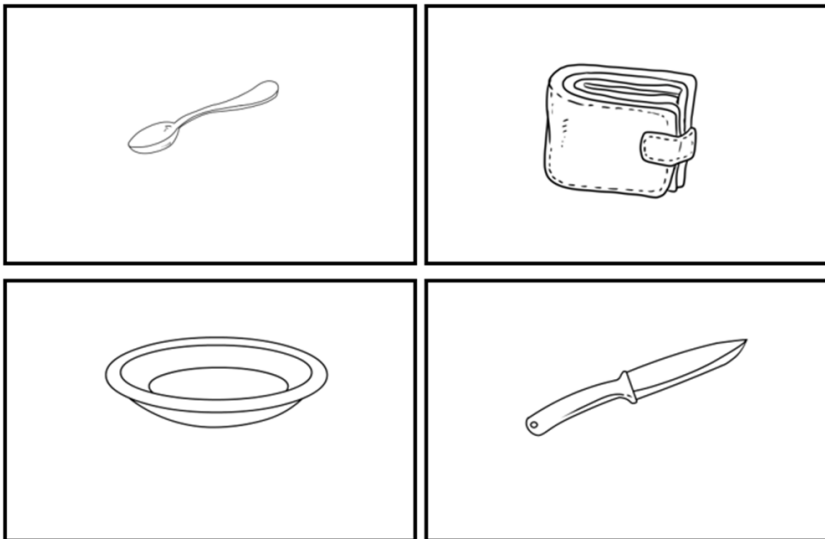
6 of 20 target words: ضفدع (Frog)



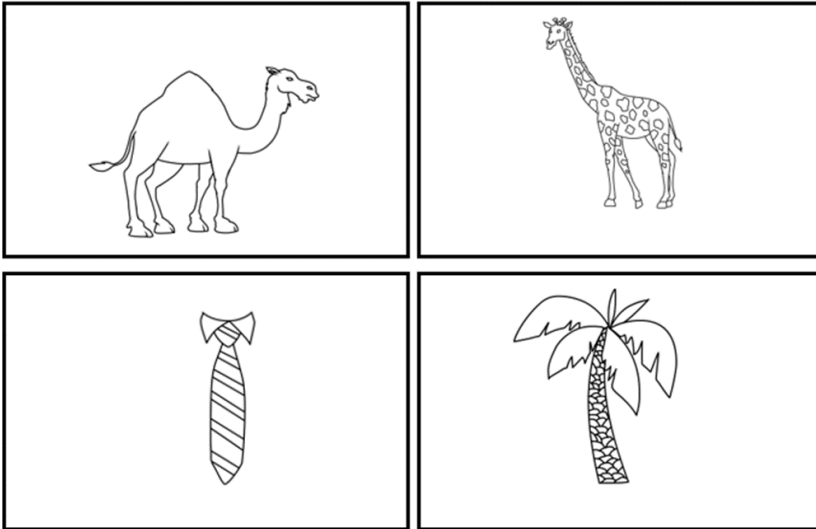
7 of 20 target words: مسبحة (Rosary)



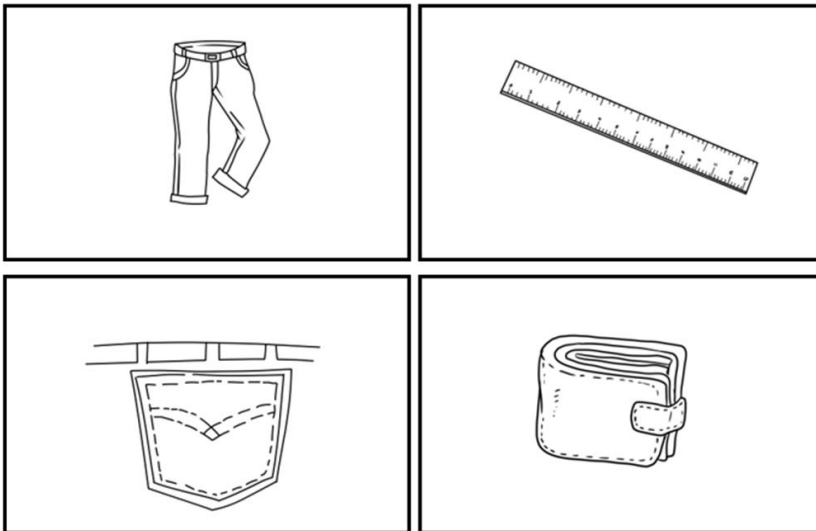
8 of 20 target words: ملعقة (Spoon)



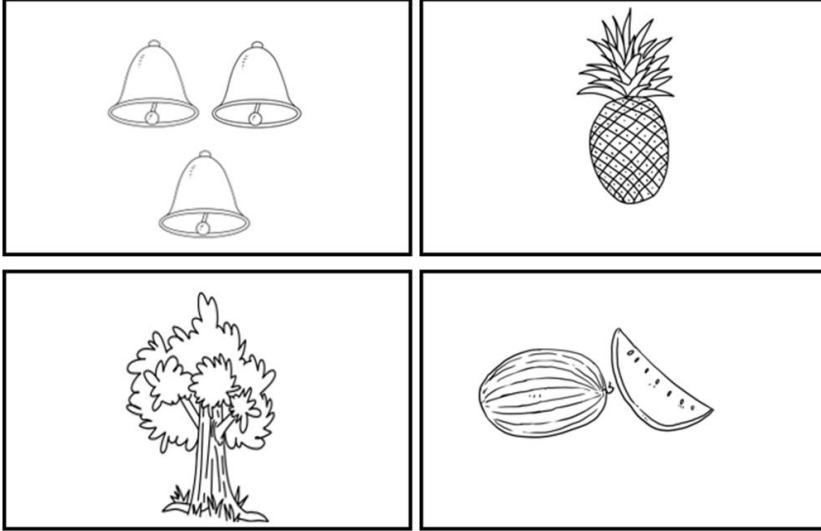
9 of 20 target words: زرافة (Giraffe)



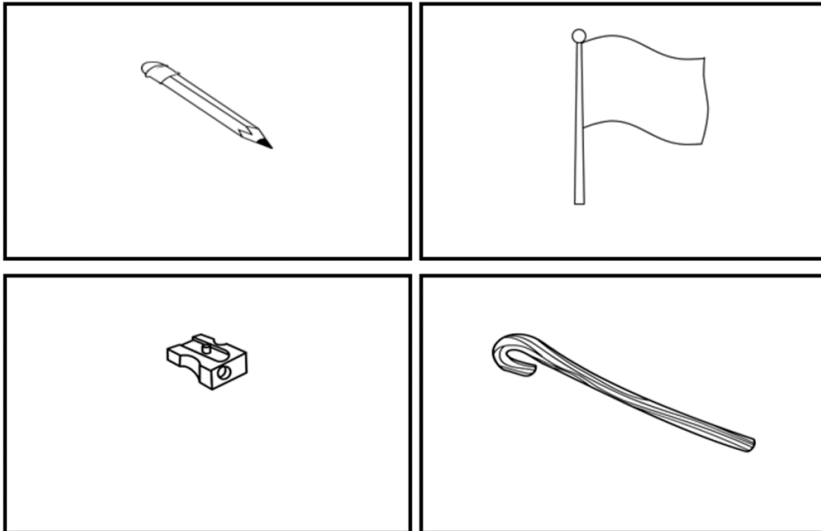
10 of 20 target words: محفظة (Wallet)



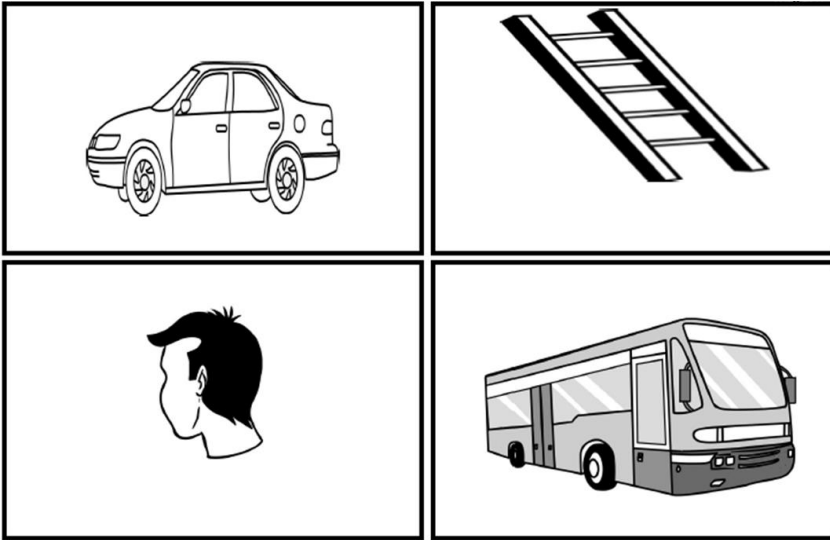
11 of 20 target words: أناناس (Pineapple)



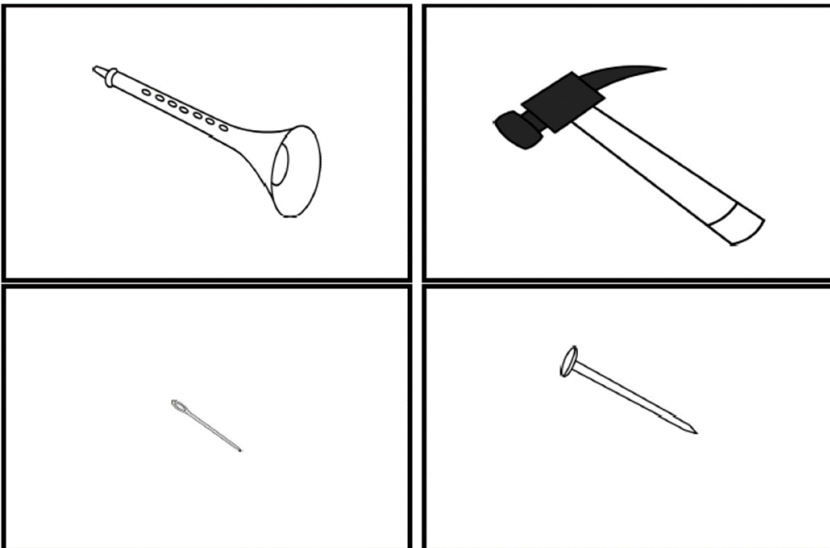
12 of 20 target words: قلم (Pencil)



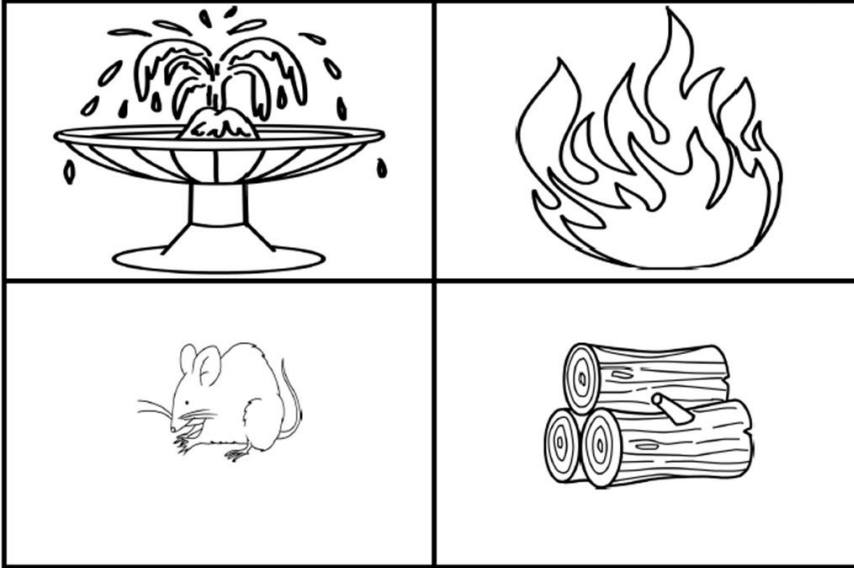
13 of 20 target words: راس (Head)



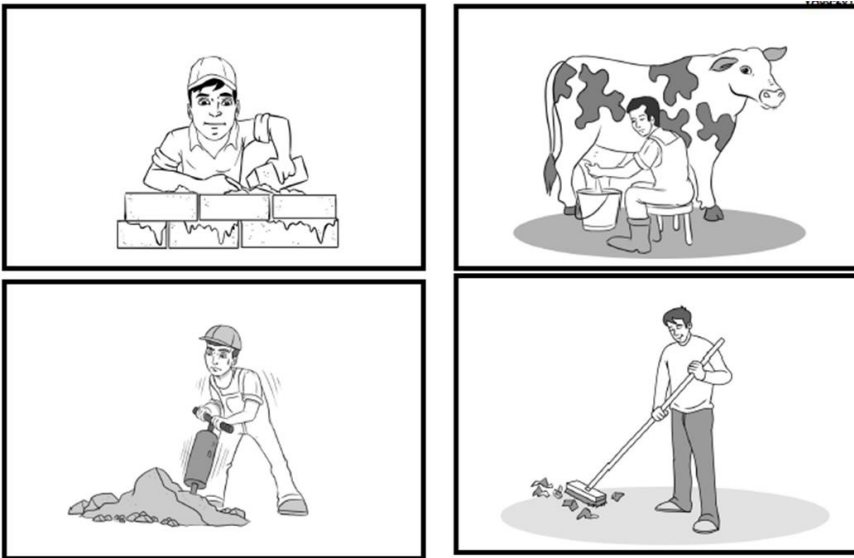
14 of 20 target words: مسمار (Nail)



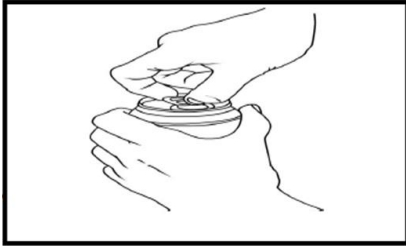
15 of 20 target words: نار (Fire)



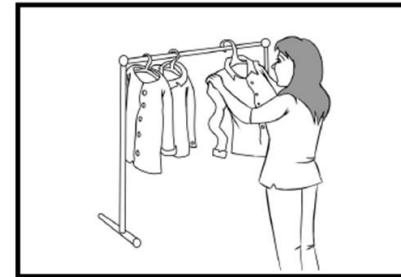
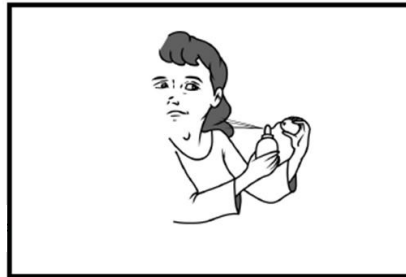
16 of 20 target words: يکنس (Sweeping)



17 of 20 target words: تاكل (Eating)



18 of 20 target words: تعلق (Hanging)



19 of 20 target words: يسوق (Driving)



20 of 20 target words: يعك (Counting)



Subtest 3 of 4: Word Comprehension

Instruction: Say the following to the examinee: "I will now read a sentence and show you four pictures. Point to the picture that represents the sentence I said."

Present the target picture with the foils making sure that all four are within the examinee's visual field.

One point is given for each correct response.

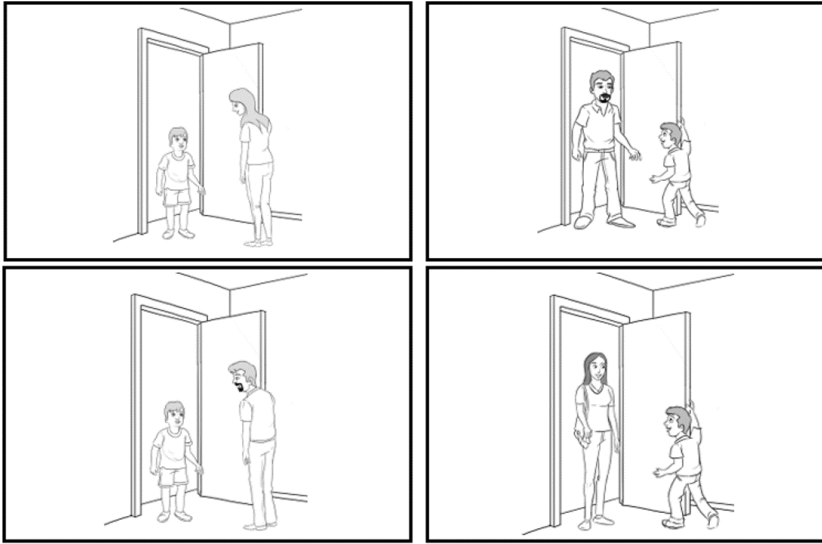
1 of 5 target sentences: البنت بتشرب مي (*The girl is drinking water*)



2 of 5 target sentences: أولاد بيكتبوا على الدفاتر (*The boys are writing on their books*)



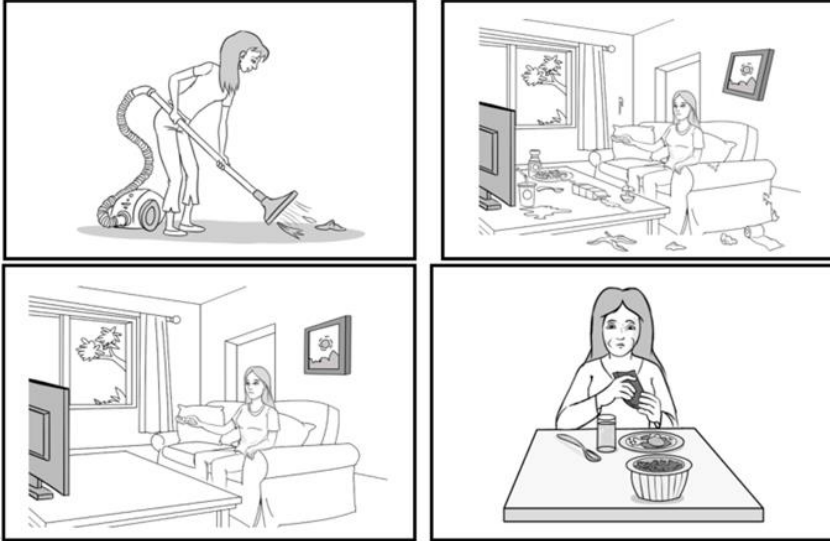
3 of 5 target sentences: الولد يفتح الباب لأبوه (The boy opened the door for his father)



4 of 5 target sentences: الطباخ بيلحق وراء الدكتور (The chef is chasing the doctor)



5 of 5 target sentences: بعد ما نظفت البيت، قعدت الست تتفرج على التلفزيون (After the woman finished cleaning the house, she sat and watched TV)



Score: _____

Subtest 4 of 4: Comprehension of Multiple Sentence and Paragraph

Read the stories to the client, and then ask the questions that follow. Circle the patient's response. Start with practice item:

أعطى المعلم للطالب المجتهد جائزة عبارة عن كتاب. هل الجائزة كانت فلوس؟

The teacher gave the good student a book as a prize. Was the student given money as a prize?

Story 1 (one sentence):

طلبت الأم من ابنها أنه يشتري حليب و هو راجع من الجامعة و لما كان الإبن مروح بسيارته, نسي يروح البقالة.

The Mother told her son to buy some milk when he was going home from college. While the son was driving home, he forgot to go to the grocery store.

1. هل كان الأبن مروح على البيت من شغله؟

Was the son driving home from work?

2. هل تذكر الإبن أن يشتري الحليب؟

Did the son remember to buy milk?

Story 2 (two sentence):

سعيد خُصص الثانوية و صار عمره 18 سنة. راح سعيد دائرة الترخيص عشان يفحص سواقة. نجح سعيد بالفحص و قالوله أنه بيقتدر ياخذ الرخصة. لقي سعيد شباك تسليم الرخص مسكر, لان دوام الموظفين في الدائرة كان مخلص.

Saed finished high school and he is now 18 years old. Saed went to the DMV to have a driving test. Saed passed and he was told that he can have his license. He later found that the office that delivers tickets was closed because working hours were finished.

1. هل سعيد عمره 16 سنة؟

Was Saed 16 years old?

2. هل نجح سعيد في فحص القيادة؟

Did Saed pass the driving test?

3. هل كان شباك تسليم الرخص مفتوح؟

Was the ticket office opened?

4. هل حصل سعيد على رخصة القيادة ذلك اليوم؟

Did Saed get his driver's license?

Story 3 (short paragraph):

Read the story to the examinee, then ask the questions that follow. Circle the patient's response.

زار صاحب المزرعة بستان الرمان. طلب من المزارع اللي بيشتغل عنده أنه يجيبه رمانة حلوة الطعم. وفي كل مرة كان يرجع فيها لصاحب المزرعة, كان المزارع يجيبه رمانة حامضة. فتعجب صاحب البستان و قال: كيف ما بتعرف مكان الرمان حلو الطعم وإلك سنتين بتشتغل في مزرعتي؟ قاله المزارع: عشان هذا البستان مو ملكي, أنا ما باكل من الرمان, فما بعرف وين الرمان الحلو طعمه. أعجب صاحب البستان بأمانة المزارع و عشان هيك, أعطاه نصف محصول الرمان.

The farm's owner visited his pomegranate field. He asked the farmer he hired to bring him a sweet pomegranate. Each time the farmer comes back with a pomegranate and gives it to the owner, the pomegranate he brought tasted sour. The owner was baffled and said: "How do you not know where the sweet tasting pomegranates are now that worked for me for two years?" The farmer said: "Because I don't own this field, I don't eat from it, therefore, I don't know where the sweet pomegranates are. The owner admired the farmer's honesty and decided to give him half the harvest of pomegranate that year.

1. هل طلب صاحب البستان رمانا حامض الطعم؟
Did the owner of the farm ask for sour pomegranates?
2. هل طلب صاحب البستان رمانا حلو الطعم؟
Did the owner of the farm as for sweet pomegranates?
3. هل تذوق صاحب البستان رمانة واحدة؟
Did the owner of the farm taste one pomegranate?
4. هل وهب صاحب البستان الحارس نصف المحصول؟
Did the owner of the farm give the famer half the harvest?
5. هل كان حارس البستان امينا؟
Was the farmer of the field an honest man?
6. هل خان الحارس الأمانة؟
Did the farmer betray his duties?

Score: _____

Part 3 of 4
Repetition Part

Present the following words and sentences verbally one at a time and ask the examinee to repeat them.

Make sure that each item is heard by examinee, repeat if required.

Use scores 1 for correct response, and 0 for incorrect or no response

- | | | |
|-----|-----------------------|----------------------|
| 1. | Monosyllabic nonword | شاس |
| 2. | Bi-syllabic nonword | حازي |
| 3. | Multisyllabic nonword | باتيكاع |
| 4. | Monosyllabic | باب |
| 5. | Monosyllabic | دال |
| 6. | Bi-syllabic | هذا |
| 7. | Bi-syllabic | شوكة |
| 8. | Bi-syllabic | برواز |
| 9. | Bi-syllabic | حصان |
| 10. | Multisyllabic | اربعة |
| 11. | Multisyllabic | اسماعيل |
| 12. | Multisyllabic | متماسك |
| 13. | Multisyllabic | أربعتش |
| 14. | word/sentence | صب القهوة |
| 15. | word/sentence | الولد وصل البيت |
| 16. | word/sentence | الرز محطوط على النار |

17. word/sentence تطير الطيارة فوق السحاب بدون عناء
18. word/sentence هيئة المواصفات و المقاييس الأردنية موجودة في منطقة خلداء

Score: _____

Part 4 of 4
Naming Part

Subtest 1 of 4: Object naming

Present the items and/or object one at a time and ask the examinee to name them.
Make sure that each item is in the patient's visual field. Record the error types following the key below.

Use scores **1** for correct response, and **0** for incorrect or no response

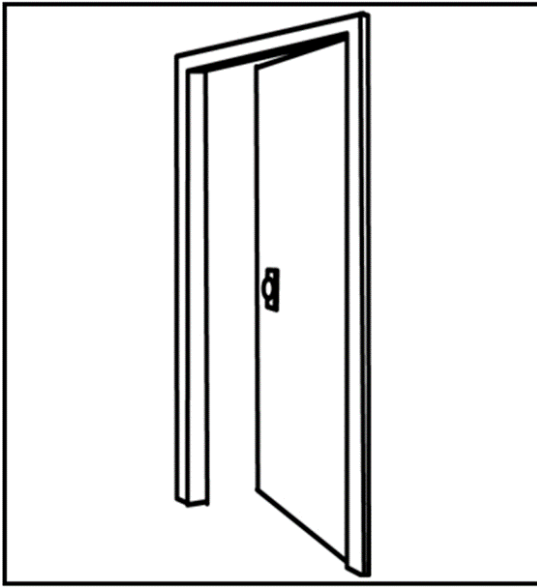
1. قلم (Pen)
2. ساعة (Wrist watch)
3. كتاب/دفتر (Book)
4. ورقة (Paper)
5. محرمة (Hygiene tissue)

Score: _____

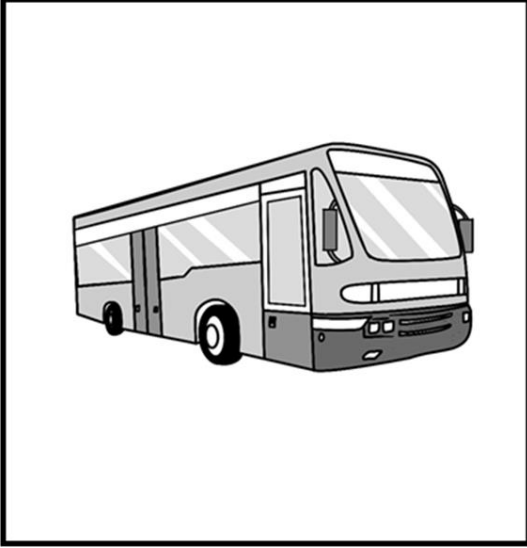
Subtest 2 of 4: Picture Naming

Present the items and/or pictures one at a time and ask the patient to name them. Make sure that each item is in the patient's visual field. Record the error types following the key below.

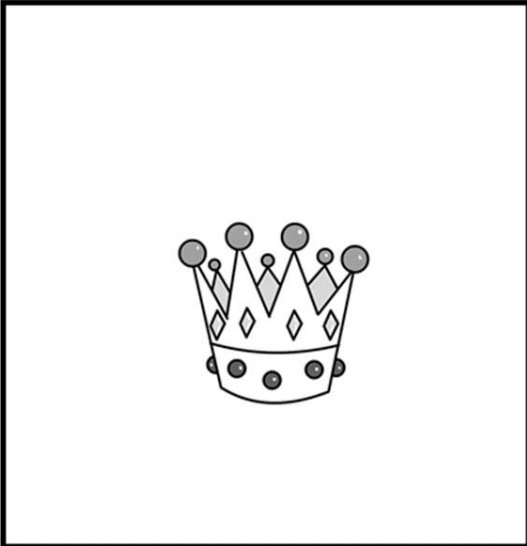
Use scores 2 for correct response, 1 for incomplete response, and 0 for incorrect or no response

1 of 15 target words: باب (Door)**2 of 15 target words: عين (Eye)**

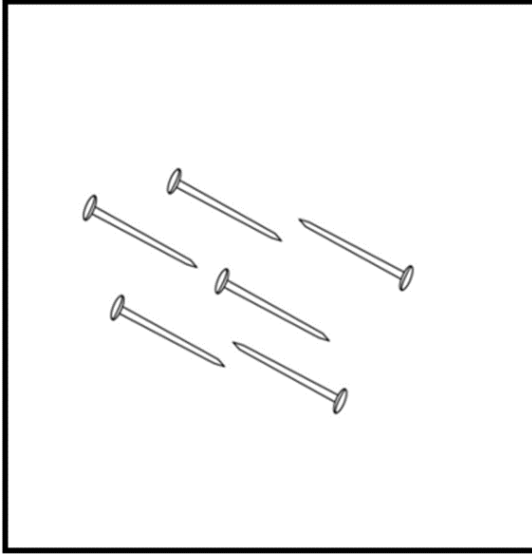
3 of 15 target words: باص (Bus)



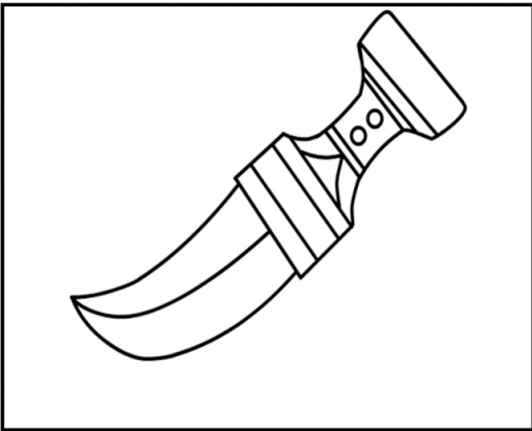
4 of 15 target words: تاج (Crown)



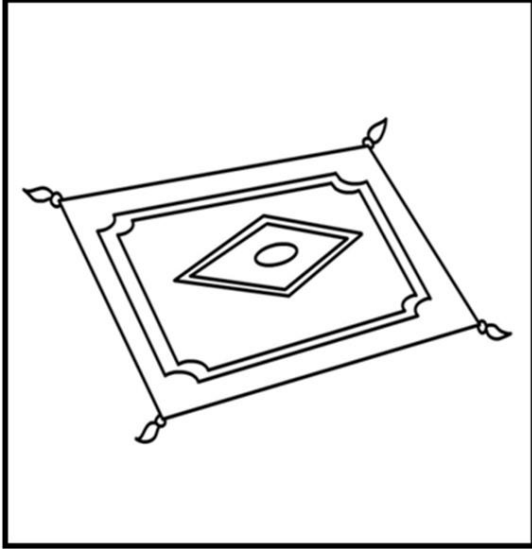
5 of 15 target words: مسامير (Nails)



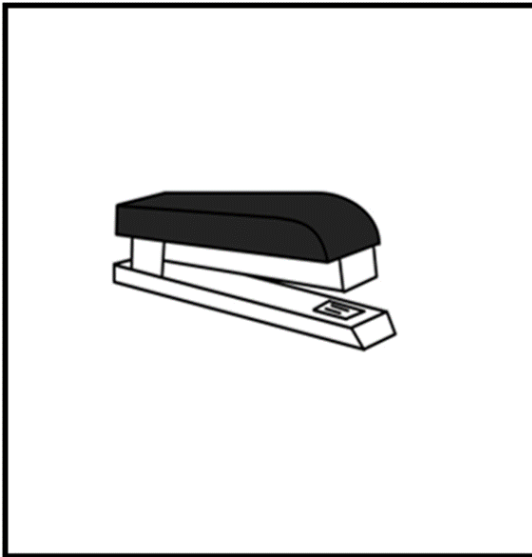
6 of 15 target words: خنجر (Dagger)



7 of 15 target words: سجادة (Carpet)



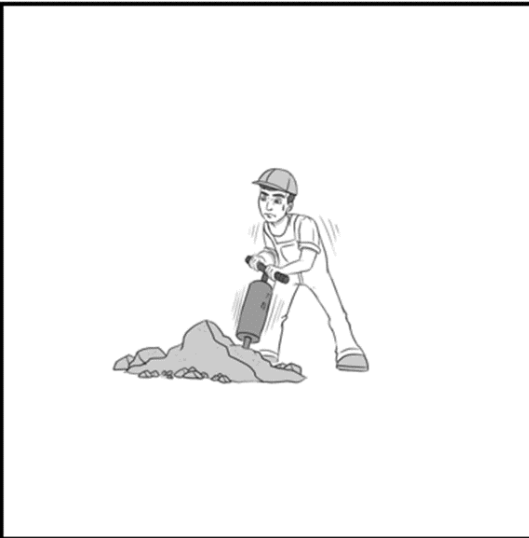
8 of 15 target words: مكبس (Stapler)



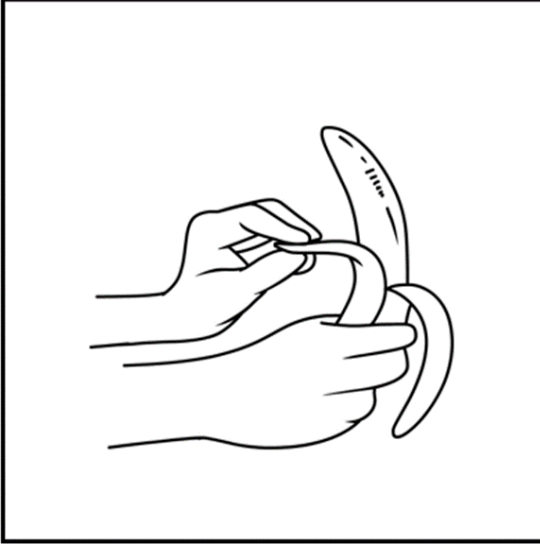
9 of 15 target words: يسوق (Driving)



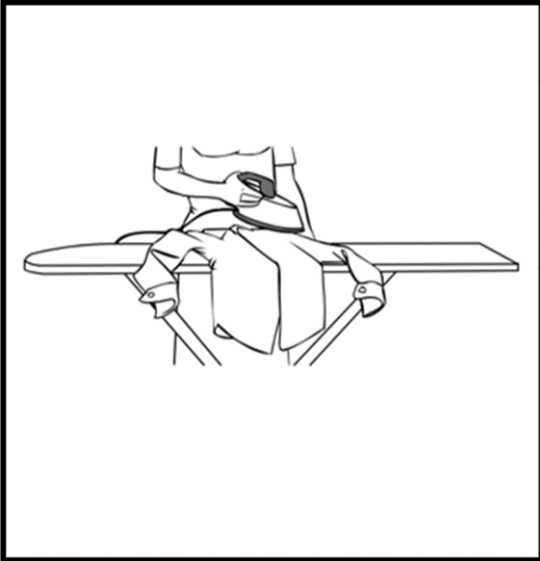
10 of 15 target words: يحفر (Digging)



11 of 15 target words: تقشر (To peel)



12 of 15 target words: تگوي (To iron)



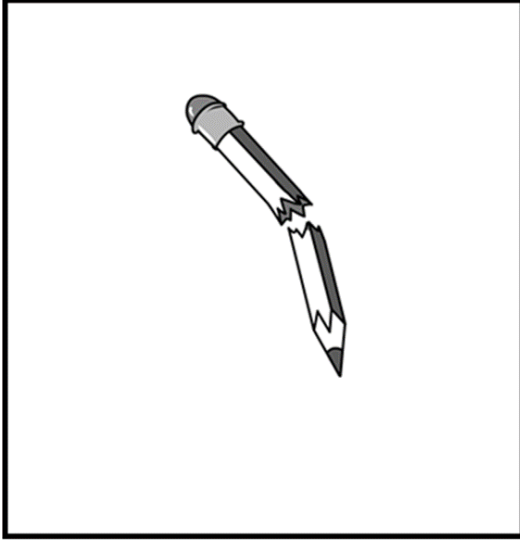
13 of 15 target words: يعض (To bite)



14 of 15 target words: نائم (Sleeping)



15 of 15 target words: مكسور (Broken)



Score: _____

Subtest 3 of 4: Responsive naming

Present the Examinee with the questions presented below asking him to provide only one-word responses.

Use scores 2 for correct response, 1 for incomplete response, and 0 for incorrect or no response

وين يروح المريض؟

Where does a sick person go? Practice item

1. (قلم) *What do we write with?* اش نستخدّم في الكتابة ؟

2. (تلاجية) *Where do we put the milk so it won't spoil?* وين نخط الحليب عشان ما يخرّب ؟

3. (مدرسة) *Where does a teacher work?* وين بيشتغل المدرس ؟

4. (يصلي) *What does a man do in a mosque?* ايش بيسوي الرجال في المسجد؟

5. (يسرق) *What does a thief do?* ايش يسوي الحرامي ؟

Score: _____

Subtest 3 of 4: Responsive naming

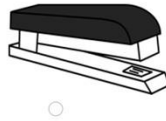
Instruction: Ask the examinee to name as many animals as he/she can in one minute. Start with the “fruits” as practice item.

Time response and record each unrepeated correct response

Number of words per minute: _____

APPENDIX B**SAMPLE ITEM IN QUESTIONNAIRE 1 OF EXPERIMENT ONE**

Sample item in Experiment One (Questionnaire 1: Verification of the JAAT-1 visual stimuli by NBD Jordanian Arabic speakers)



What is the word that best describe the above picture?

- Ruler
- Stapler
- Sharpener
- I do not know

APPENDIX C

QUESTIONNAIRE 2 ITEMS APPRAISING THE JAAT AS A UNIT

The Eight Questionnaire Items in Questionnaire 2 of Experiment Three Appraising the JAAT as a Unit

1. How would you describe administration time of the test? Response options: “too long, too short, and suitably timed”
2. Overall, do you consider this test linguistically and culturally sensitive for Jordanian persons with aphasia? Response options: “yes, no, not sure” In the case of the participant answering the question with “no” or “not sure”, the participant is asked to elaborate more in a designated item in Questionnaire 2.
3. The test can provide a diagnostic description of receptive and expressive language in persons with aphasia. Response options: “strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, and strongly disagree”
4. The test can determine the severity levels within persons with aphasia. Response options: “strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, and strongly disagree”
5. The test can help in designing a treatment plan for persons with aphasia. Response options: “strongly agree, agree, somewhat agree, neither agree nor disagree, somewhat disagree, disagree, and strongly disagree”
6. Would this test be useful in identifying other impairments associated with aphasia such as apraxia and dysarthria? Response options: “no, yes, and not sure” In the case of the participant answering the question with “no” or “not sure”, the participant is asked to elaborate more in a designated item in Questionnaire 2.
7. If this test was available to you for use right now, how likely would you be to use it? Response options: “likely, somewhat likely, not sure, somewhat unlikely, and unlikely”
8. If this test is available for you now, how would you rank it based on potential use? Response portions: “First choice, second choice, not sure, and will not use” In the case of the participant answering the question with “no” or “not sure,” the participant is asked to elaborate more in a designated item in Questionnaire 2.

APPENDIX D**SAMPLE OF SUBTEST APPRAISING QUESTIONNAIRE ITEMS**

Sample of the Subtest Appraising Item in Experiment Three (Questionnaire 2: Validation of Test Content Based on SMEs Ratings of the JAAT-3) Showing the Four Standard Subtest Statements



Above is a sample from the Picture description subtest. This subtest assesses expository speech.

Please select your level of agreement on the following statements.

APPENDIX E**SAMPLE OF PART APPRAISING QUESTIONNAIRE ITEM**

Sample of a Part Appraising Item in Experiment Three (Questionnaire 2: Validation of Test Content Based on SMEs Ratings of the JAAT-3) Showing the Standard Question

Part one: Verbal Fluency Question

When considering all the three subtests of the Verbal Fluency Section,

Does this part contain **enough items** to assess verbal fluency in Jordanian persons with aphasia?

- Yes
- No
- Not Sure

If your answer was no or not sure, please write your reasons in the space below:

APPENDIX F**QUESTIONNAIRE ITEMS APPRAISING THE JAAT AS A UNIT**

The Final Eight Questionnaire Items in Experiment Three (Questionnaire 2: Validation of test Content Based on SMEs Ratings of the JAAT-3) that Address the Appraisal of the JAAT as a Unit

This section's items will target your professional impression of the test as a complete unit

Item 1: How would you describe administration time of the test?

- Too Short
- Too Long
- Suitably Timed

Item 2: Do you consider this test linguistically and culturally sensitive for Jordanian persons with aphasia?

- Yes
- No
- Not Sure

If your answer was no or not sure, please write your reasons in the space below

Considering diagnosing aphasia symptoms, please rate the test using the scales below regarding each sentence:

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
Item 3: The test can provide a diagnostic description of receptive and expressive language in persons with aphasia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Item 4: The test can determine severity levels within persons with aphasia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Item 5: The test can help in designing a treatment plan for persons with aphasia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Item 6: Would this test be useful in identifying other impairments associated with aphasia such as apraxia and dysarthria?

- Yes
- No
- Not Sure

If your answer was no or not sure, please write your reasons in the space below

Item 7: If this test were available to you for use right now, how likely would you be to use it?

- Very likely
- Somewhat likely
- Likely
- Somewhat unlikely
- Unlikely

Item 8: If this test is available for you now, how would you rank it based on potential use?

- First choice
- Second choice
- I would not use it
- Not Sure

If your answer was “I would not use it” or “not sure,” please write your reasons in the space below

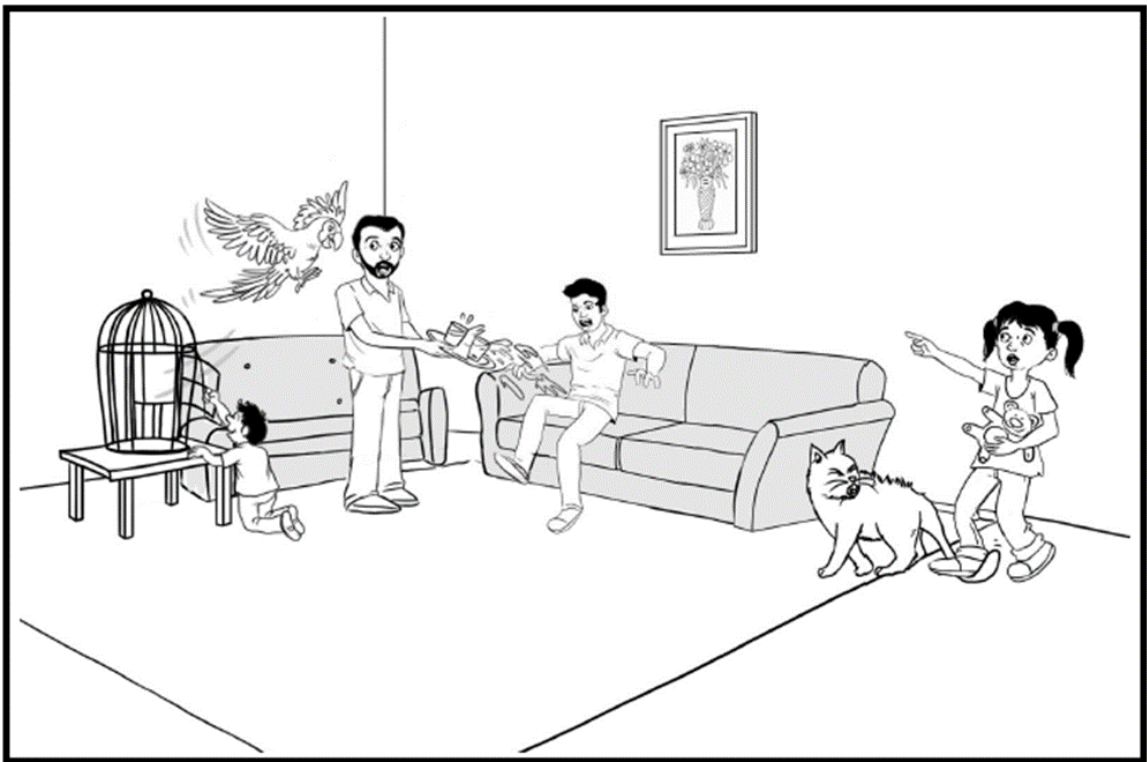
APPENDIX G

MODIFIED OR ADDED ITEMS TO JAAT AFTER EXPERIMENT TWO

Part 1 of 3

Changed or Alternative Items Used in JAAT-2 after the Results of Experiment Two

Picture description subtest in Verbal Fluency Part



Part 2 of 3
Alternative Items Used in JAAT-2 after Results of Experiment Two

Picture Sequence Description Subtest in Verbal Fluency Part with Added Items on Table



Part 3 of 3
Alternative items used in JAAT-2 after the results of Experiment Two

Picture naming subtest items in Naming Part replacing items *carpet*, *stapler*, *to iron*, and *to bite*

