

Cite this Article: Khan, M., Butt, A. K., & Anjum, F. A. (2026). Development and Validation of an Indigenous Urdu Naming Test for Assessing Paraphasia in Adults With Aphasia. *Journal of Regional Studies Review*, 5(4), 49-56. <https://doi.org/10.62843/jrsr/v.iv.224>

RESEARCH ARTICLE

JOURNAL OF REGIONAL STUDIES REVIEW (JRSR)

Development and Validation of an Indigenous Urdu Naming Test for Assessing Paraphasia in Adults With Aphasia

Maroof Khan ^a Ayesha Kamal Butt ^b Fayyaz Ahmed Anjum ^c

Corresponding Author: Fayyaz Ahmed Anjum; fayyaz.anjum@myu.edu.pk



Abstract: Paraphasia, a disruption of word-level language production that commonly follows stroke, traumatic brain injury, or other neurological damage, is frequently assessed with confrontation naming instruments, yet very few such tools have been developed for Urdu, the national language of Pakistan and a first or second language for several hundred million speakers across South Asia. This two-phase study developed and validated an indigenous Urdu Naming Test (UNT) for identifying and classifying paraphasic errors in adults with aphasia. In Phase I, an item pool of 50 picture-naming stimuli was generated from open-ended interviews with 20 neurologically healthy adults and 50 patients with aphasia, refined through review by three subject-matter experts, and then administered to 230 patients with aphasia (ages 22-80) drawn from hospitals in Rawalpindi and Islamabad. Exploratory factor analysis using principal component extraction with direct oblimin rotation supported a three-factor, 30-item structure accounting for a substantial share of the total variance, and the scale demonstrated strong internal consistency (Cronbach's alpha = .89). In Phase II, construct validity was examined in an independent sample of 100 patients with aphasia (ages 20-45) by correlating UNT scores with the Boston Naming Test (BNT) and the Wechsler Intelligence Scale for Children (WISC). The UNT correlated strongly and positively with the BNT ($r = .83$), supporting convergent validity, and correlated strongly and negatively with the WISC ($r = -.69$), supporting discriminant validity. These findings suggest that the UNT is a reliable and valid, culturally and linguistically appropriate instrument for assessing paraphasia among Urdu-speaking patients with aphasia, with implications for clinical assessment and rehabilitation planning in the region.

Keywords: Aphasia, Anomia, Paraphasia, Urdu Naming Test, Confrontation Naming, Factor Analysis, Construct Validity, Pakistan

Introduction

Aphasia is an acquired disorder of language caused by injury or disease to the brain areas involved in language processing, which affects the understanding and expression of spoken and written language without being associated with a general intellectual or sensory-motor deficit (Goodglass et al., 1983; Murray & Clark, 2006). Aphasia is often accompanied by a common and debilitating symptom called anomia, or word-finding difficulty, which is often characterised by paraphasia, or the use of incorrect, invented, or only partially related words in place of the intended ones (Murray & Clark, 2006). The evaluation of paraphasic errors is a fundamental clinical task, as naming is a central component in functional communication, for clinicians working with stroke, TBI, or progressive neurological disease patients.

^a Senior Lecturer, Sindh Institute of Physical Medicine and Rehabilitation, Karachi, Sindh, Pakistan.

^b Professor/Head of Department, Speech and Language Department, Riphah University, Islamabad, Pakistan.

^c Associate Professor, Department of Psychology, Muslim Youth University, Islamabad, Pakistan.

Although the clinical and research focus on aphasia has increased significantly, the majority of confrontation-naming tests have been created and normed on English-speaking populations. This is very restrictive because it is known that the performance obtained in a language assessment tool is influenced by the frequency of names used in the language, the phonological structure of the names, the familiarity of the language community with the objects depicted on the tools, and other factors, which are not universal and will differ from one language community to another (de la Plata et al., 2008). Urdu is the national language of Pakistan and is used in one form or another by several hundred million people in Pakistan, India, Bangladesh and Nepal and has close structural links with Hindi (Wallraff, 2000). Such was the reach, however, that clinicians evaluating Urdu-speaking aphasic patients have had to use translated or adapted Western instruments that may not fully reflect the phonological and lexical patterns unique to the Urdu language and have been limited in terms of cultural competence in assessment and intervention (Plata, 2008).

Errors often fall into one of the following overlapping categories: Paraphasia in the form of phonemic (or literal) errors refers to errors in the pronunciation of the sound or syllable, such as saying 'louse' instead of 'ice' (Laganaro et al., 2009). Semantic (or verbal) paraphasia is a case where the substituted word is semantically and meaningfully related to the target, like saying "television" for "computer. Paraphasia is the distortion of a word into a less familiar form; in neologistic paraphasia, less than 50% of the intended word is produced correctly; sometimes an invented, unrecognisable form of the word is produced. Other subtypes are perseverative paraphasia, when a previous word response intrudes; remote paraphasia, when the substituted word is only remotely related to the intended word; multi-word paraphasia, when the same error occurs across several words; and off-target utterances, which have no obvious relationship to the intended word. Other subtypes are perseverative paraphasia, in which a previous 'response' to a word intrudes; remote paraphasia, when the substituted word is only remotely related to the intended word; and multi-word paraphasia, in which the same error occurs across several words; off-target utterances have no obvious relationship to the intended word. Studies of the structure of anomie errors in other languages, like German (Kulke & Blanken, 2001), Hebrew (Gollan & Frost, 2001), Italian and Spanish (Kiran & Thompson, 2003), indicate that it is itself sensitive to the grammatical architecture of the speaker's language.

Conventionally, naming tests require participants to name pictures instead of words, partly because a large percentage of the population in some areas (like Pakistan) is not literate, so pictures are more generally usable (de la Plata et al., 2008). An additional difficulty for clinicians is that anatomical and functional imaging studies suggest that there may be some variation in the representation of each of the language systems of a bilingual or multilingual speaker in the language network in the perisylvian region (Gomez-Tortosa et al., 1995), further emphasising the necessity of language-specific rather than translated assessment tools. In addition to the clinical relevance of using these naming tests to aid in the diagnosis of stroke-related aphasia (Murray & Clark, 2006), they have been used in the diagnosis of early language changes related to Alzheimer's disease and other progressive disorders.

The present study was conducted to design and validate an Urdu Naming Test (UNT), a new instrument specially developed for the recognition and quantification of paraphasia in Urdu-speaking adults who are aphasic due to the huge Urdu-speaking population and the lack of any psychometrically validated instrument for naming for this population. Two phases of study were conducted. Phase I involved the development of an item pool and subsequent testing of the internal factor structure and reliability of the new instrument. Phase II examined the construct validity of the UNT by testing its relationship with an established naming measure and with a measure of general cognitive ability. Specifically, the study addressed the following objectives: (a) to develop an indigenous, culturally appropriate measure of paraphasia for use with Urdu-speaking patients with aphasia; (b) to establish the internal consistency reliability of the new measure; and (c) to establish the construct validity of the UNT, including its convergent validity in relation to the Boston Naming Test (BNT) and its discriminant validity in relation to a measure of general cognitive ability, the Wechsler Intelligence Scale for Children (WISC).

Method

Research Design

The study used a two-phase instrument-development design combining qualitative item generation with quantitative psychometric evaluation. Phase I comprised item generation through interviews, expert content review, exploratory factor analysis, and reliability estimation. Phase II independently examined the convergent and discriminant validity of the resulting scale. Each phase drew on an independent sample, consistent with recommended practice in scale-development research.

Phase I: Development of the Urdu Naming Test

Item Generation: An item pool was generated using semi-structured, picture-based interviews conducted in Urdu. Fifty line drawings of common nouns were prepared, and participants were asked to name each pictured object within a fixed response window. Interviews were conducted one-on-one and were about 40 to 50 minutes. Each participant's responses were recorded verbatim by interviewers, and written informed consent was obtained from all patients or, where appropriate, their carers before data collection started.

The sample consisted of 20 neurologically healthy adults and 50 adults with a clinical diagnosis of aphasia (ages 22-80 years) from the Armed Forces Institute of Rehabilitation Medicine and Combined Military Hospital / Military Hospital Rawalpindi. Responses were transcribed and analyzed using a deductive content-analysis approach (Hsieh & Shannon, 2005; Mayring, 2000), in which responses were coded into predefined categories corresponding to the paraphasia subtypes described above (phonemic, semantic, neologistic, perseverative, remote, multi-word, and off-target). Response patterns that occurred with sufficient frequency across participants were retained and used to draft the initial pool of 50 picture-naming items, each of which could be scored for the presence and type of paraphasic error.

Expert Review: The 50-item pool was reviewed by a three-member expert panel comprising two psychologists holding doctoral degrees and one physician certified in rehabilitation medicine. The panel evaluated each item for clarity, redundancy, cultural appropriateness, and correspondence with the target paraphasia subtypes, and suggested wording or presentation changes where needed. Following this review, items were retained in a picture-based format accompanied by written cues in both Urdu and English, together with a semantic cue and a phonemic cue to be used in scoring partial responses. All pictorial stimuli used in the pool were freely available images sourced for research use.

Scale-development Sample and Procedure: The 50-item pool was subsequently administered to an independent sample of 230 patients with aphasia, aged 22 to 80 years, recruited from hospitals in the twin cities of Rawalpindi and Islamabad; the majority of participants were male, reflecting the composition of the clinical population from which the sample was drawn. Indoor and outpatient patients were approached with the assistance of hospital staff, and each participant (or caregiver, where applicable) provided informed consent before testing. Responses to each item were scored across seven indices: the number of spontaneously correct responses, the number of stimulus cues provided, the number of correct responses following a stimulus cue, the number of phonemic cues provided, the number of correct responses following a phonemic cue, the number of multiple-choice prompts provided, and the number of correct multiple-choice responses. A response made spontaneously with no error (within 30 seconds) received two points; a response with an error made after the 30 seconds elapsed received one point; and responses with impaired articulation or paraphasia were recorded in an extra error-code column.

Statistical Analysis: The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity (Hair et al., 1994; Tabachnick & Fidell, 2007) were used to determine the sample's adequacy for factor analysis. Item-total correlations were computed to confirm that all items related meaningfully to the overall scale. Because the paraphasia subtypes were expected to be conceptually related rather than fully independent, factors were extracted using principal component analysis with direct oblimin (oblique) rotation, which is recommended over orthogonal rotation methods such as varimax when underlying

constructs are likely to be correlated (Field, 2005). Internal consistency reliability was subsequently estimated using Cronbach's alpha.

Phase II: Validation of the Urdu Naming Test

Participants: An independent sample of 100 patients with aphasia, aged 20 to 45 years (91 men; M = 28.88, SD = 8.17), was recruited for the validation phase. All participants had experienced a stroke or traumatic brain injury resulting in aphasia and were undergoing rehabilitation at the time of testing. Educational attainment ranged from primary schooling to postgraduate study.

Instruments: Three instruments were administered: the 50-item Urdu Naming Test developed in Phase I; the Boston Naming Test (BNT; Kaplan et al., 2001), a 60-item, difficulty-graded picture-naming instrument widely used to assess word-retrieval impairment in aphasia and related conditions; and the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), used here as a measure of general cognitive ability against which the discriminant validity of the UNT could be evaluated.

Procedure: Participants were tested individually. After providing informed consent, each participant received standardized written instructions for the UNT, the BNT, and the WISC before each instrument was administered. Basic demographic information was collected on a separate form, and confidentiality of responses was assured. Cronbach's alpha was again computed for each instrument within this sample, and Pearson product-moment correlations were computed between UNT scores and scores on the BNT and the WISC to evaluate convergent and discriminant validity, respectively.

Ethical Considerations

All procedures involving human participants were conducted in accordance with standard ethical principles for research with clinical populations. All participants (and those unable to give informed consent by themselves) gave written informed consent before the collection of any data. All participants were informed of the purpose of the research, and their responses were kept confidential throughout the research. The authors declare no conflict of interest.

Results

Sampling Adequacy and Item Analysis

Prior to factor extraction, the adequacy of the 230-participant sample was evaluated. The Kaiser-Meyer-Olken measure of sampling adequacy was .70, which is within the acceptable-to-good range for factor analysis set by Kaiser (1974) and is above the .50 criterion. The test of sphericity for the inter-correlation matrix used on Bartlett's test was statistically significant ($p < .001$), indicating that the matrix was appropriate for conducting factor analysis. Individual item-total correlations ranged from .30 to .83 ($p < .05$ to $p < .01$), which showed meaningful correlations both among the items and with the total score on the scale.

Table 1

Kaiser-Meyer-Olkin Measure and Bartlett's Test of Sphericity for the Urdu Naming Test Item Pool (N = 230)

Kaiser-Meyer-Olkin Measure (KMO)	Bartlett's Test of Sphericity	df	p
.70	Significant	—	< .001

Note. Exact chi-square and degrees-of-freedom values from the original dataset could not be reproduced with certainty and are therefore omitted; the significance level is reported as recorded.

Factor Structure

The item pool did not contain orthogonal subsyndromes, as there was a conceptual relationship among them, so principal component analysis (direct oblimin rotation) was used. Four factors were initially extracted, having eigenvalues greater than 1.0. Factor I accounted for 19.94% of the total variance (eigenvalue = 7.37), Factor II accounted for 19.65% (eigenvalue = 7.27), Factor III accounted for 11.51% (eigenvalue = 4.25)

and Factor IV accounted for 9.41% (eigenvalue = 3.48) of variance, with all four factors explaining 60.51% of variance (Table 2).

Table 2

Eigenvalues and Percentage of Variance Explained by the Four Extracted Factors (N = 230)

Factor	Eigenvalue	% of Variance	Cumulative %
I	7.37	19.94	19.94
II	7.27	19.65	39.59
III	4.25	11.51	51.10
IV	3.48	9.41	60.51

The fourth factor was not kept in the final scale. The items in this factor loaded only weakly (around .39) and were otherwise not conceptually coherent with each other or with the other three factors, indicating that this factor did not indicate a stable and interpretable dimension of paraphasia. In the end, 30 items were retained across three oblique factors, with each factor loading higher than the traditional .38 cut-off for retention. The factor matrix for the retained items is presented in Table 3.

Table 3

Factor Matrix of the 30 Retained Items of the Urdu Naming Test (N = 230)

Factor I (15 items)	Loading	Factor II (9 items)	Loading	Factor III (6 items)	Loading
Item 2	.537	Item 23	.585	Item 1	.688
Item 4	.829	Item 29	.948	Item 3	.430
Item 7	.536	Item 30	.823	Item 5	.768
Item 8	.510	Item 31	.755	Item 16	.429
Item 10	.629	Item 32	.802	Item 25	.505
Item 11	.829	Item 33	.601	Item 26	.576
Item 12	.601	Item 34	.852		
Item 14	.865	Item 35	.914		
Item 15	.514	Item 37	.526		
Item 18	.761				
Item 19	.716				
Item 20	.422				
Item 21	.574				
Item 22	.395				
Item 24	.422				

Note. Loadings shown are from the principal component solution with direct oblimin rotation.

On the basis of the content of the items loading on each factor, Factor I was interpreted as reflecting more pervasive, global naming failure across a broad range of stimuli; Factor II was interpreted as reflecting a cluster of more specific, targeted retrieval failures; and Factor III was interpreted as reflecting a smaller cluster of related error patterns. These labels are descriptive rather than definitive, and their theoretical interpretation would benefit from replication in future research.

Reliability

Internal consistency reliability of the 50-item UNT, computed on the Phase I sample of 230 patients, was high (Cronbach's alpha = .89), indicating strong internal consistency for the pool as a whole. As shown in Table 4, reliability estimates computed independently in the Phase II validation sample (N = 100) were likewise satisfactory for all three instruments administered.

Table 4

Internal Consistency Reliability of the Urdu Naming Test, Boston Naming Test, and Wechsler Intelligence Scale for Children (N = 100, Phase II)

Measure	Number of Items	Cronbach's Alpha
Urdu Naming Test (UNT)	50	.89
Boston Naming Test (BNT)	60	.78
Wechsler Intelligence Scale for Children (WISC)	50	.84

Convergent and Discriminant Validity

Construct validity of the UNT was examined by correlating UNT scores with scores on the BNT (expected to be conceptually similar, supporting convergent validity) and the WISC (expected to be conceptually distinct, supporting discriminant validity). As shown in Table 5, UNT scores correlated strongly and positively with BNT scores, $r = .83$, indicating that patients who exhibited more paraphasic errors on the UNT also tended to show greater naming impairment on the BNT. In contrast, UNT scores correlated strongly and negatively with WISC scores, $r = -.69$, indicating that greater paraphasia on the UNT was associated with lower general cognitive-ability scores, while remaining a conceptually distinct construct rather than a direct proxy for cognitive ability. The correlation between the BNT and the WISC was comparatively weak, $r = -.23$, further supporting the discriminant separation of naming impairment from general cognitive functioning.

Table 5

Inter-Scale Correlation Matrix for the Urdu Naming Test, Boston Naming Test, and Wechsler Intelligence Scale for Children (N = 100, Phase II)

Measure	BNT	WISC
UNT	.83	-.69
WISC	-.23	—

Note. UNT = Urdu Naming Test; BNT = Boston Naming Test; WISC = Wechsler Intelligence Scale for Children.

Discussion

The present study set out to develop and validate an indigenous naming instrument for the assessment of paraphasia among Urdu-speaking adults with aphasia. Results from Phase I revealed a pool of 50 items; the result of interviews and expert refinement described a psychometrically coherent three-factor, 30-item structure with high internal consistency. This supports the feasibility of building an assessment tool for paraphasia from the ground up in a language and cultural context where few, if any, comparable instruments previously existed, echoing calls in the broader literature for population- and language-specific normative instruments rather than direct translations of Western measures (de la Plata et al., 2008).

The retention of a three-factor rather than a single-factor or four-factor structure is broadly consistent with clinical descriptions of paraphasia as a multidimensional phenomenon encompassing phonemic, semantic, and more idiosyncratic (neologistic, perseverative, or off-target) error types (Laganaro et al., 2009; Murray & Clark, 2006). The exclusion of a fourth, weakly loading factor suggests that not every candidate dimension identified at the interview stage translated into a stable, statistically distinct component, which is a common outcome in early-stage scale development and points to a useful direction for refinement in subsequent revisions of the UNT.

The validity findings from Phase II provide encouraging, if preliminary, support for the UNT as a measure of naming impairment. The strong positive correlation between the UNT and the well-established BNT ($r = .83$) suggests that the two instruments are capturing substantially overlapping aspects of confrontation-naming ability, which is the pattern one would expect if the UNT is validly measuring paraphasia. At the same time, the strong negative correlation with the WISC ($r = -.69$) indicates that UNT performance is not

simply a proxy for general cognitive ability; rather, patients with more severe naming impairment also tended to show lower performance on a task tapping broader cognitive functioning, a pattern consistent with the well-documented co-occurrence of language and broader cognitive deficits following stroke and traumatic brain injury. The comparatively weaker BNT-WISC correlation ($r = -.23$) further supports the interpretation that naming impairment and general cognitive ability, while related, are separable constructs.

These results should be interpreted in light of several limitations. First, although the overall sample across both phases was reasonably large, patients presenting with clinically significant paraphasia are relatively uncommon and geographically dispersed within the health system from which participants were drawn, which constrained recruitment and may limit the representativeness of the sample relative to the broader population of Urdu-speaking patients with aphasia. Secondly, the study did not have a non-aphasic comparison group for the purpose of directly testing the discriminative power of the UNT against non-aphasics. Third, the conceptual framework used to organize item generation was adapted for the purposes of this study, and the underlying theoretical model of paraphasia remains an area of active debate, with some researchers emphasizing memory-based accounts and others emphasizing broader cognitive-linguistic accounts; future revisions of the UNT would benefit from closer alignment with a single, well-specified theoretical model of lexical access and retrieval (Kiran & Thompson, 2003). Finally, because the UNT was administered as a clinician-rated, behaviorally scored task rather than a self-report measure, its psychometric properties should be interpreted specifically within that administration format.

Conclusion

This study developed and provided initial validation for the Urdu Naming Test, an indigenous picture-naming instrument designed to assess paraphasia among Urdu-speaking adults with aphasia. Across two independent samples, the UNT demonstrated a coherent three-factor structure, strong internal consistency, and construct validity evidenced by a strong positive correlation with the Boston Naming Test and a strong negative correlation with the Wechsler Intelligence Scale for Children. Based on these findings, it is concluded that the UNT is a promising culturally appropriate instrument in the limited battery of existing naming assessment tools for the Urdu-speaking clinical population and could be used in naming assessment and for monitoring change throughout the process of speech and language rehabilitation. Longitudinal designs should be used in future studies to provide predictive and test-retest reliability, non-aphasic comparison groups to enhance the discriminant validity evidence, and larger and more geographically varied samples of Urdu-speaking patients to replicate the work. Further development (theoretical underpinning of its factors) would add to the utility of the UNT as a normed clinical and research tool for the evaluation of paraphasia in the region.

References

- De La Plata, C. M., Vicioso, B., Hynan, L., Evans, H. M., Diaz-Arrastia, R., Lacritz, L., & Cullum, C. M. (2008). Development of the Texas Spanish Naming Test: A test for Spanish speakers. *The Clinical Neuropsychologist*, 22(2), 288–304. <https://doi.org/10.1080/13854040701250470>
- Field, A. P. (2005). *Discovering statistics using SPSS* (2nd ed.). Sage Publications.
- Gollan, T. H., & Frost, R. (2001). Two routes to grammatical gender: Evidence from Hebrew. *Journal of Psycholinguistic Research*, 30(6), 627–651. <https://doi.org/10.1023/A:1014235223566>
- Gomez-Tortosa, E., Martin, E. M., Gaviria, M., Charbel, F., & Ausman, J. I. (1995). Selective deficit of one language in a bilingual patient following surgery in the left perisylvian area. *Brain and Language*, 48(3), 320–325. <https://doi.org/10.1006/brln.1995.1014>
- Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *Boston diagnostic aphasia examination*. Lea & Febiger.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1994). *Multivariate data analysis* (3rd ed.). Macmillan.
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *The Boston Naming Test* (2nd ed.). Lippincott Williams & Wilkins.
- Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research*, 46(4), 773–787. [https://doi.org/10.1044/1092-4388\(2003/061\)](https://doi.org/10.1044/1092-4388(2003/061))
- Kulke, F., & Blanken, G. (2001). Phonological and syntactic influences on semantic misnamings in aphasia. *Aphasiology*, 15(1), 3–15. <https://doi.org/10.1080/02687040042000070>
- Laganaro, M., Morand, S., & Schnider, A. (2009). Time course of evoked-potential changes in different forms of anomia in aphasia. *Journal of Cognitive Neuroscience*, 21(8), 1499–1510. <https://doi.org/10.1162/jocn.2009.21117>
- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research*, 1(2), Article 20. <https://doi.org/10.17169/fqs-1.2.1089>
- Murray, L. L., & Clark, H. M. (2006). *Neurogenic disorders of language: Theory driven clinical practice*. Thomson Delmar Learning.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon.
- Wallraff, B. (2000, November). What global language? *The Atlantic Monthly*, 286(5), 52–61.
- Wechsler, D. (1949). *Wechsler intelligence scale for children*. Psychological Corporation.