


Detection of Major and Mild Neurocognitive Disorder with the Verbal Naming Test

Kelsey Spalding-Wilson^{1,*} , Grant Harris¹, Victoria A. Windham², Adriana M. Strutt² and Brian P. Yochim¹

¹VA Saint Louis Health Care System, Saint Louis, MO, USA

²Neuropsychology Section, Department of Neurology, Baylor College of Medicine, Houston, TX, USA

(RECEIVED December 30, 2020; FINAL REVISION June 7, 2021; ACCEPTED June 15, 2021; FIRST PUBLISHED ONLINE August 26, 2021)

Abstract

Objective: The Verbal Naming Test (VNT) is an auditory-based measure of naming or word finding. The current multisite study sought to evaluate the reliability and validity of the VNT in the detection of major and mild neurocognitive disorder (NCD). **Method:** This study analyzed clinical data from two outpatient neuropsychology clinics ($N = 188$ and $N = 77$) and a geriatric primary care clinic ($N = 104$). Cronbach's alpha and Spearman correlations with other measures were calculated. ROC analyses were used to calculate sensitivity, specificity, positive predictive power, and negative predictive power for the detection of major and mild NCD per DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) criteria. **Results:** The VNT was found to have strong reliability (Cronbach's alpha = .90) and high convergent validity with a commonly used picture-naming task (NAB Naming, Spearman's $r_{ho} = .65$, $p < .001$). The VNT showed good sensitivity and specificity for the detection of NCDs, particularly major NCD, with an area under the curve of .85, sensitivity of .80, and specificity of .75. A possible discontinue rule is also suggested for clinicians to use. **Conclusions:** These findings provide compelling evidence for the use of the VNT to detect neurocognitive impairment in a clinical setting. The VNT provides a reliable alternative to picture-naming tasks, which may be advantageous when working with visually impaired patients or conducting evaluations over telehealth.

Keywords: Word-finding, Naming, Dementia, MCI, Assessment, Language

INTRODUCTION

The assessment of confrontation naming, or word finding, is regularly included as part of a neuropsychological assessment battery. A critical component of language functioning, impaired naming can result from damage to the left hemisphere (e.g., due to stroke, traumatic brain injury, cancer, or temporal lobe epilepsy) and is characteristic of nearly all forms of aphasia (Blumenfeld, 2010; Lezak et al., 2012). In addition, impaired naming is included in the diagnostic guidelines for dementia or major neurocognitive disorder (NCD), as defined by both the National Institute on Aging (NIA)/Alzheimer's Association guidelines (McKhann et al., 2011) and the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; American Psychiatric Association, 2013). Performance on confrontation naming tasks has been shown to distinguish Alzheimer's disease from other causes of dementia, such as Huntington's disease or vascular disease,

and from other conditions such as major depressive disorder (Braaten et al., 2006; Salmon & Bondi, 2009) and therefore can play a critical role in differential diagnosis. Given the rising incidence of dementia, the assessment of naming remains a critical piece of the neuropsychological evaluation.

Historically, naming has been assessed using picture-naming tasks in which participants are shown a series of line drawings (e.g., the Boston Naming Test) or images (e.g., the Neuropsychological Assessment Battery [NAB] Naming subtest) of objects and asked to provide the name of each object. A major limitation of such measures is that they rely on visual abilities to test a verbal function, so visual impairment can interfere with the validity of these measures. Low vision or blindness is estimated to affect over 3.8 million Americans over the age of 45 (Chan et al., 2018). Beyond vision loss associated with aging, visual deficits can also be associated with acquired brain injury or neurodegenerative conditions, leading to difficulty using visual measures to assess other aspects of cognition. Therefore, it is advantageous to have an alternative modality to assess naming, such as an auditory-based measure.

*Correspondence and reprint requests to: Kelsey Spalding-Wilson, PhD, VA St. Louis Health Care System, 1 Jefferson Barracks Drive, St. Louis, MO 63125, USA. E-mail: kelsey.wilson@va.gov

Additionally, nonvisual measures may be preferable for tele-neuropsychological evaluations (Brearly et al., 2017). This testing modality can provide increased access to services for individuals who may have significant travel constraints or geographic limitations. Recently, psychologists have seen a significant increase in the use of telehealth services in the context of the global COVID-19 pandemic. The use of video conferencing for neuropsychological evaluation allows for the provision of services while simultaneously minimizing the risk of disease transmission. Even if tested in person, non-visual measures minimize the need for sharing of stimulus materials between patient and examiner.

Auditory-based naming measures, or naming to definition tasks, have been proposed as a possible alternative to picture-naming tasks. One such task is the Verbal Naming Test (VNT) (Yochim et al., 2015), in which the patient is given a series of verbal prompts, and the patient is asked to provide the word described. The VNT was developed in a sample of older adults, and word frequency was taken into consideration, such that items are ordered from highest to lowest frequency of use (ranging from 5.29 to .55 per million words) and are used more rarely in conversation than items on other naming tasks (Yochim et al., 2013). The use of lower frequency words may increase its sensitivity to emerging word finding difficulty (Goodglass et al., 2001). The VNT was shown to have strong internal consistency (Cronbach's $\alpha = .84$) and was sensitive to the detection of dysnomia (Yochim et al., 2015). Since its initial publication as a 55-item test, descriptive data from a sample of healthy older adults have been made available to use in neuropsychological evaluations (Wynn et al., 2020). These data are based on a revised 50-item version of the scale, which eliminated a few items found to be problematic in clinical practice (e.g., items that prompted multiple responses that may be considered correct). Wynn et al. (2020) additionally provided compelling evidence for the feasibility of the VNT for tele-neuropsychological assessment. In this preliminary investigation, performance on the VNT when administered in person was significantly correlated with VNT performance one week later when administered over the phone (Wynn et al., 2020), suggesting that this measure may be of use for tele-neuropsychological evaluations.

These initial studies provide strong evidence that the VNT may serve as a feasible alternative to commonly used picture-naming tasks. The present study sought to further examine the clinical utility of the VNT by investigating its use in a larger clinical sample. While initial clinical data are based on a 55-item version of the measure (Yochim et al., 2015), the present study used the revised 50-item version, for which descriptive data were developed (Wynn et al., 2020). In the initial validation study, evidence suggested that the VNT is sensitive to the detection of dysnomia (Yochim et al., 2015). The current study sought to 1) expand upon this finding by examining the sensitivity and specificity of the VNT in detecting major and mild NCD, 2) explore relationships between the VNT and a number of demographic variables in a different geographic sample, 3) use empirical data to

develop a discontinue rule, which could shorten the administration time, particularly for patients experiencing significant difficulty on the test, and increase the feasibility of the measure, and 4) provide an exploration of the correlations between the VNT and other neuropsychological measures.

METHOD

Participants

This study was conducted with approval from the Institutional Review Boards at VA Saint Louis Health Care System and Baylor College of Medicine and in accordance with the Helsinki Declaration. Data were obtained from three clinical samples who presented for neuropsychological assessment to determine the presence and severity of a NCD. Sample 1 included 188 patients from an outpatient neuropsychological assessment clinic at the VA Saint Louis Health Care System in Saint Louis, Missouri. Sample 2 included 104 patients referred for evaluation of dementia in an outpatient geriatric primary care clinic at the VA Saint Louis Health Care System. Sample 3 included 77 patients from a Neuropsychology Clinic at Baylor College of Medicine in Houston, Texas. To increase the generalizability of our findings as well as increase power, these samples were combined for analyses into one large sample ($N = 369$).

Demographic information for the total sample, as well as each individual sample, is presented in Table 1. Participants in Sample 2 were significantly older than those in Samples 1 and 3, $t(290) = -8.09$, $p < .001$, and $t(179) = 6.43$, $p < .001$, respectively, but there was not a significant difference in age between Sample 1 and Sample 3, $p > .05$. Participants in Sample 3 had a significantly higher level of education than participants in Samples 1 and 2, $t(263) = -5.22$, $p < .001$, and $t(179) = -6.10$, $p < .001$, respectively. However, there was no significant difference in education between Sample 1 and Sample 2, $p > .05$. Additionally, participants in Sample 2 obtained significantly lower VNT scores than those in Sample 1, $t(290) = 3.18$, $p < .01$. However, VNT scores for Sample 3 did not differ significantly from Sample 1 or Sample 2, $p > .05$, for each. Racial distribution also differed significantly across the three groups, such that Sample 1 had a higher proportion of Black participants than Samples 2 and 3 ($\chi^2 = 21.41$, $p < .01$). For the purposes of this study, the three samples were combined to increase power and generalizability of the results.

All patients were tested face-to-face, in person, before the COVID-19 pandemic began. Exclusion criteria included age less than 40 or over age 89 (as ages over 89 are considered protected health information), and evidence of insufficient effort as determined by free-standing measures of test-taking effort when available. All diagnoses were made by a board-certified neuropsychologist or geropsychologist using DSM-5 criteria for major and mild NCD. To prevent circularity, given that the VNT was a focus of study, clinical diagnoses were not based on the VNT. Demographic information for the total sample as well as each individual sample are

Table 1. Demographics and additional sample information

	Total Sample (N = 369)	Sample 1 (N = 188)	Sample 2 (N = 104)	Sample 3 (N = 77)
Age	69.52 (9.5)	66.68 (9.9)	75.44 (6.6)	68.47 (8.0)
Sex	316 M, 53 F	175 M, 13 F	101 M, 3 F	40 M, 37 F
Education (years)	13.50 (2.8)	13.27 (2.6)	12.67 (2.6)	15.19 (2.9)
Race/Ethnicity				
White	280 (76%)	128 (68%)	90 (87%)	62 (81%)
Black	76 (21%)	57 (30%)	14 (14%)	5 (6%)
Hispanic	8 (2%)	2 (1%)	0 (0%)	6 (8%)
Asian American	4 (1%)	1 (.5%)	0 (0%)	3 (4%)
Native American	1 (.3%)	0 (0%)	0 (0%)	1 (1%)
No Diagnosis	136 (37%)	80 (43%)	31 (30%)	25 (32%)
Mild NCD	108 (29%)	51 (27%)	37 (36%)	20 (26%)
Major NCD	121 (33%)	55 (29%)	34 (33%)	32 (42%)
Unspecified NCD	4 (1%)	2 (1%)	2 (2%)	0 (0%)
VNT Score	43.39 (6.4)	44.26 (5.7)	41.78 (7.4)	43.44 (6.3)

Note. Numbers corresponding to age, education, and VNT scores reflect sample means and standard deviations. Numbers corresponding to Race/Ethnicity, and Diagnoses indicate the number and percent of patients in each category. M = male, F = female, NCD = neurocognitive disorder.

available in Table 1. Raw scores from the measures below were used in data analyses.

Measures

All of the participants were administered the VNT. However, as the current study used data obtained from clinical samples, not all participants were administered the same test battery. Therefore, there is variation in the supplemental tests used for exploratory analyses of convergent and divergent validity. For example, some participants were administered the California Verbal Learning Test – Second Edition (CVLT-II) as a measure of verbal memory, while others may have been administered the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) List Learning Test. All measures included in our analyses are described here.

Verbal Naming Test

On the VNT (Yochim et al., 2015), patients are given a verbal description of a noun or verb and asked to provide the corresponding word (e.g. “What is the name of the thing you hold over your head when it rains?” [i.e., *umbrella*]). Patients have 10 seconds to give the correct response before a phonemic cue is provided indicating the sound of the first syllable (e.g. “it starts with ‘um-.’”). The patient then has an additional 10 seconds to provide a response; however, a response provided after the phonemic cue does not count toward the total correct score. The original VNT consisted of 55 items, listed in Yochim et al. (2015). However, descriptive data from a large sample were made available for a revised 50-item version, which eliminated 5 items found to be problematic in clinical practice (Wynn et al., 2020). Only responses to the items retained in the revised 50-item version were used for analyses. Therefore, possible scores ranged from 0 to 50 with higher scores indicating better naming performance.

NAB Naming Subtest (NAB Naming)

On the NAB Naming test (Stern & White, 2003), patients are shown an image and asked to name the object presented. Scores range from 0 to 31 and reflect the number of items named correctly without a cue.

California Verbal Learning Test-Second Edition

The CVLT-II (Delis et al., 2000) is a word list-learning task. The current study used the total words learned across Trials 1–5 (Trials 1–5 Total; scores range from 0 to 80) and the number of words recalled after a 20 minute delay (Long Delay Free Recall; scores range from 0 to 16).

Letter Fluency

Patients were given one minute to verbally generate as many words as possible that begin with the same letter. In this study, letter fluency refers to raw scores obtained from both the Verbal Fluency subtest of the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001) and the FAS subtest of the Controlled Oral Word Association Test (Benton & Hamsher, 1976; Spreen & Benton, 1969), which uses the same letter prompts.

D-KEFS Category Fluency

On the Category Fluency portion of the Verbal Fluency subtest of the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001), patients are given one minute to verbally generate as many words as possible that belong to the same category. This is administered for two distinct categories in two separate trials.

WAIS-IV

Select subtests from the Wechsler Adult Intelligence Scale-Fourth Edition (Wechsler, 2008) were included. *Block Design* is a measure of visuoconstructional ability in which patients are asked to recreate designs using red and white blocks. Total scores range from 0 to 66. *Visual Puzzles* is a visuospatial task in which participants must mentally piece together three puzzle pieces to reconstruct a presented image. Total scores range from 0 to 26. *Coding* is a measure of psychomotor processing speed in which patients are given 120 seconds and asked to rapidly record symbols that correspond to a set of digits. Total scores range from 0 to 135.

Repeatable Battery for the Assessment of Neuropsychological Status

The RBANS is a 12-subtest cognitive screening battery (Randolph, 1998, 2012). Eight subtests were used in this study. *List Learning* is a test of rote verbal learning while *Story Memory* evaluates contextualized verbal learning. *List Recall* and *List Recognition* represent delayed retrieval of verbal information from the *List Learning* subtest in a free-recall and recognition format, respectively. *Story Recall* tests for delayed free recall of information presented during *Story Memory*. *Figure Copy* involves drawing a complex geometric figure, examining visuospatial reasoning and visuomotor skills. *Figure Recall* involves free recall of the design shown during *Figure Copy* and examines visual memory. *Line Orientation* tests basic visuospatial skills by requiring the examinee to correctly identify spatial orientation of two lines. *Picture Naming* tests confrontation naming skills when presented with visual stimuli. *Semantic Fluency* tests language retrieval when given a categorical prompt.

Trail Making Test – Trial A

Trails A is a timed test of visual attention/processing. Patients are instructed to connect a set of 25 circles/dots numerically as quickly as possible via line drawing. A maximum of 180 seconds is allowed.

Analyses

Statistical analyses were performed with SPSS-25.

Demographic Variables and VNT Performance

Relationships between the VNT and demographic variables were explored using Spearman correlations, due to its non-normal distribution, and *t*-tests. Analyses examining race were limited to patients who identified as Black/African American or White/European American, as this captured 99% of the sample ($N = 289$).

Reliability

Cronbach's alpha was used to assess the internal consistency of the VNT. This analysis was done using only data from Samples 1 and 2 ($N = 292$) as item-level data were not available for Sample 3.

Detection of Major and Mild Neurocognitive Disorder

A one-way ANOVA and follow up *t*-tests were used to determine whether VNT performance differed significantly between individuals who received no cognitive diagnosis, a diagnosis of major NCD or a diagnosis of mild NCD. ROC analyses were conducted to assess how well the VNT detected major NCD or mild NCD. Sensitivity, specificity, positive predictive power, and negative predictive power were calculated.

Development of a Discontinue Rule

ROC analyses were used to determine the score which reliably differentiated between mild NCD and major NCD with 95% specificity. This score will be used as a proposed cutoff score, indicating that the test can be discontinued.

Preliminary Investigation of Convergent and Divergent Validity

Spearman's rho (r_s) correlations were used to examine relationships between VNT performance with performance on other neuropsychological tests. As data were obtained from clinical samples, not all participants were administered the same test battery (e.g., not all participants were administered the CVLT-II). Therefore, validity analyses do not include the entire sample of participants. The number of participants included in each analysis is included in the results section. To determine whether the VNT correlated more strongly with measures hypothesized to reflect convergent validity (NAB Naming and Category Fluency) than with measures hypothesized to reflect discriminant validity (Block Design and Visual Puzzles), correlations were compared using Fisher's *r* to *z* transformation (Steiger, 1980). Given the variability in sample size, these analyses are considered preliminary in nature.

RESULTS

Demographic Variables and VNT Performance

In the total sample, 121 patients received a diagnosis of major NCD, 108 patients received a diagnosis of mild NCD, and 136 patients did not receive a cognitive diagnosis. Four patients were diagnosed with unspecified NCD and were excluded from analyses that involved diagnosis as a variable of interest. Diagnostic and demographic information for the total sample, as well as each individual sample, is presented in

Table 1. Overall, VNT performance significantly correlated with age, $r_s = -.41$, $p < .001$, and education, $r_s = .21$, $p < .001$. To see whether this result was unique to the VNT, the same analyses were run with the NAB Naming test. NAB naming performance also significantly correlated with age, $r_s = -.25$, $p < .001$, and education, $r_s = -.17$, $p < .01$, although these correlations were somewhat lower. Additionally, individuals who identified as White obtained significantly higher scores on the VNT (mean [SD] = 44.16 [5.7]) than those who identified as Black (mean [SD] = 41.21 [7.6]), $t(99) = 3.15$, $p < .01$. The same was true for the NAB naming test, $t(77) = 4.18$, $p < .001$ (Mean [SD] NAB Score for White participants = 29.29 [2.3], and for Black participants = 27.25 [3.6]). Additional analyses indicated that individuals who identified as White had more years of education (mean = 13.70 years) than individuals who identified as Black (mean = 12.75 years), $t(354) = 2.65$, $p < .01$. In a limited sample ($N = 76$), the VNT also correlated with an estimate of premorbid functioning (Test of Premorbid Functioning), $r_s = .28$, $p < .05$. Although variation in gender was limited, the present study did not find evidence for a difference in VNT performance between men and women, $t(367) = -.28$, $p = .777$. The same was true when investigating potential group differences using only Sample 3, in which there was relatively equal representation of men and women, $t(75) = .23$, $p = .820$. In a limited sample ($N = 76$), the VNT was not significantly related to depressive symptoms (as measured by the Geriatric Depression Scale, a self-report measure), $r_s = .05$, $p = .689$.

Reliability

Cronbach's alpha for the VNT was .90, reflecting strong internal reliability.

Detection of Major and Mild Neurocognitive Disorder

VNT performance differed significantly across diagnostic groups, $F(2, 362) = 64.9$, $p < .001$. Follow-up analyses revealed that individuals diagnosed with mild NCD obtained significantly lower scores on the VNT than those who were given no diagnosis, $t(171) = 5.66$, $p < .001$, $d = .87$. Likewise, individuals diagnosed with major NCD performed significantly worse on the VNT than those diagnosed with mild NCD, $t(198) = 5.77$, $p < .001$, $d = .82$ and those who were not diagnosed with a cognitive disorder, $t(149) = 10.25$, $p < .001$, $d = 1.68$ (See Figure 1).

Receiver operator characteristic (ROC) curves were used to further evaluate the ability of the VNT to distinguish between those with and without a neurocognitive disorder. These analyses were used to identify possible cut scores and subsequent sensitivity, specificity, positive predictive power, and negative predictive power for each. Much of the current literature in this area (e.g., Belleville et al., 2017; Stasenko et al., 2019; Weissberger et al., 2017) only

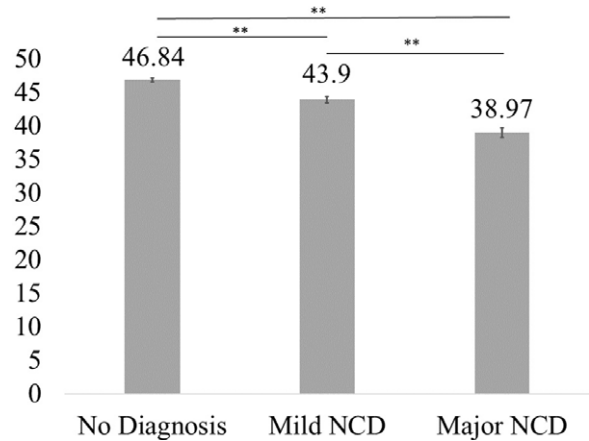


Fig. 1. Mean VNT performance across groups. Error bars represent standard deviation.

** $p < .01$.

reports sensitivity and specificity. Therefore, cut scores were selected which resulted in the closest possible sensitivity and specificity values. However, we also included alternative cut scores that result in the closest possible values of positive and negative predictive power (see Figure 2 and Table 2 for results).

Development of a Discontinue Rule

A potential discontinue rule was identified by ascertaining a score that can reliably distinguish between major and mild NCD. A total score of 34.5 has a 95% specificity for distinguishing between major and mild NCD and a positive predictive power of 88%. Therefore, if a patient has missed a total of 16 items, the test can be discontinued. If the patient were to hypothetically get every subsequent item correct after missing 16 items, their score would still be in the range of major NCD. By discontinuing after missing 16 items, the clinician can choose to spare the patient unnecessary frustration and save examination time. It is important to note that the 16 items do not have to be consecutive; if any 16 items are missed throughout the test, even if they are non-consecutive, the test can be stopped.

Validity

Spearman correlations between the VNT and other measures are found in Table 3. As shown, VNT scores correlated with the majority of tests administered. As a preliminary analysis, Fisher's r to z transformation (Steiger, 1980) was used to determine whether the VNT correlated more strongly with measures thought to represent convergent validity (NAB Naming and Category Fluency) than with measures thought to represent discriminant validity (Block Design and Visual Puzzles). Results indicated that correlations between the VNT and NAB Naming test were significantly stronger than correlations with Block Design or Visual Puzzles ($z = 2.91$, $p < .01$ and $z = 2.05$, $p < .05$, respectively). The correlation

Table 2. Receiver operating characteristic curve results

	AUC	Cut Score	Sensitivity	Specificity	Positive PP	Negative PP
<i>No Diagnosis v. Major NCD</i>	.854	44.5 45.5	.71 .80	.83 .75	.79 .74	.76 .81
<i>No Diagnosis v. Mild NCD</i>	.712	44.5 46.5	.49 .69	.83 .68	.7 .63	.67 .73
<i>Mild NCD v. Major NCD</i>	.700	43.5 44.5	.67 .71	.68 .51	.70 .62	.65 .61

Note. This table provides psychometric properties of the VNT in differentiating between the three diagnostic groups. The table includes two distinct cut scores for each comparison. The bolded rows indicate the cut score for which sensitivity and specificity were closest together. The nonbolded rows indicate the scores for which positive and negative predictive power were closest together. AUC = area under the curve, Positive PP = positive predictive power, Negative PP = negative predictive power.

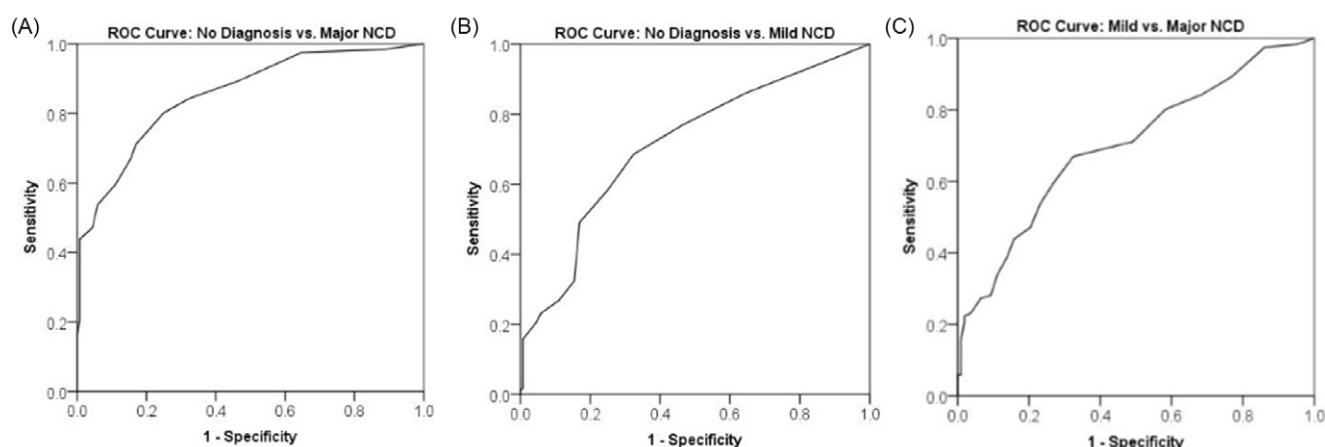


Fig. 2. Receiver Operating Characteristic Curves.

A = No Diagnosis v. Major NCD; B = No Diagnosis v. Mild NCD; C = Major NCD v. Mild NCD.

Table 3. Spearman’s Rho correlations between Verbal Naming Test, NAB naming test, and other variables

Test	VNT	N	NAB	N
NAB Naming	.653**	257	–	
Category Fluency	.589**	159	.427**	157
Letter Fluency	.558**	236	.430**	229
RBANS Semantic Fluency	.603**	103	–	
RBANS Picture Naming	.446**	101	–	
CVLT Trials 1–5	.541**	172	.465**	170
CVLT Long Delay Free Recall	.490**	171	.499**	169
RBANS List Learning	.537**	103	–	
RBANS List Recall	.310**	103	–	
RBANS Story Immediate	.513**	103	–	
RBANS Story Delayed	.525**	103	–	
Visual Puzzles	.551**	208	.399**	203
Block Design	.447**	163	.306**	161
Coding	.499**	191	.249*	187
RBANS Line Orientation	.486**	101	–	
RBANS Figure Copy	.168	101	–	
Trails A	–.602**	179	–.414**	72

Note. ** $p < .001$, * $p < .01$.

between the VNT and Category Fluency was stronger than the correlation between the VNT and Block Design ($z = 2.04$, $p < .05$), but was not significantly different from the correlation between the VNT and Visual Puzzles ($z = .69$, $p = .49$).

DISCUSSION

An assessment of naming is typically included as part of a neuropsychological assessment. This skill is typically assessed using picture-naming tasks, such as the NAB Naming Test. However, in clinical practice, it is often advantageous to use multiple methodologies when assessing a cognitive domain, such as using auditory and visual measures of memory. Auditory-based measures may also prove more clinically useful for use with patients with visual impairment or when conducting assessments via telehealth. The present study provides psychometric data on the use of the VNT in the detection of neurocognitive disorders.

Earlier work found that the VNT can be reliably used to detect dysnomia (Yochim et al., 2015). The present study found support for the use of the VNT in the detection of major and mild NCD. As expected, patients who did not receive a cognitive diagnosis performed better than those who received a diagnosis of mild NCD, and patients with mild NCD performed better than those with major NCD.

Prior work (Weissberger et al., 2017) has suggested that memory tests have a high sensitivity and specificity for detecting Alzheimer's disease, with sensitivity ranging from 71 to 93% and specificity ranging from 75 to 89%, depending on the modality of testing (e.g., a list learning task or a visual memory task). As may be expected, given the more modest level of decline observed in mild cognitive impairment (MCI), sensitivity and specificity values for detection of MCI were lower, with sensitivity ranging from 69 to 74% and specificity ranging from 74 to 82% (Weissberger et al., 2017). In comparison to memory tasks, the VNT showed comparable sensitivity and specificity for the detection of major NCD, with an AUC of .85 and sensitivity of 80% and specificity of 75%. Consistent with Weissberger and associates (2017), the VNT had lower sensitivity and specificity for the detection of mild NCD, although values were somewhat comparable to the use of memory tests (AUC of .71, 69% sensitivity, and 68% specificity). Additionally, the VNT showed a reasonable ability to distinguish between mild and major NCD (AUC of .70, 70% sensitivity, and 68% specificity). These data support the use of the VNT in neuropsychological test batteries designed to detect the presence of a neurocognitive disorder.

There is less literature examining the ability of naming tests to detect MCI or dementia. The AUC of .85 found for the VNT in detecting major NCD is comparable to the AUC of .93 found by Katsumata et al. (2015) in using the Boston Naming Test to detect dementia, and the VNT's AUC of .71 is similar to the AUC of .69 found by Katsumata in using the Boston Naming Test to detect MCI. Another picture-naming test, the Multilingual Naming Test

(MINT) was found to have an AUC of .85 for detecting dementia, and an AUC of .68 for detecting MCI (Stassenko et al., 2019).

In addition to having good predictive validity, the VNT showed strong internal reliability, with a Cronbach's alpha of .90. Previously, the VNT has been found to correlate strongly with measures of naming (NAB Naming Test) as well as other language tasks, but to show weaker correlations with tests of other neuropsychological constructs, such as visuospatial tasks (e.g. Judgment of Line Orientation) (Yochim et al., 2015). The current study sought to replicate this finding. One limitation of the current study is that given the variability in test batteries, the sample sizes for these correlations were variable, and smaller than the sample size of the study as a whole. Therefore, these analyses are considered preliminary in nature. Preliminary analyses suggest that there was a trend such that the VNT correlated more strongly with the NAB Naming Test (thought to reflect convergent validity) than measures thought to reflect discriminant validity (Block Design and Visual Puzzles). Similarly, the correlation between the VNT and Category Fluency was significantly larger than the correlation between the VNT and Block Design. Surprisingly, there was no significant difference between the correlations with Category Fluency and Visual Puzzles. Overall, while there was some preliminary evidence of convergent and divergent validity, additional work is needed to explore this area further.

One contributing factor may be that the VNT appeared to correlate with most measures administered. This raises concern that the VNT may not represent a pure measure of naming. Upon further investigation, this phenomenon did not appear to be unique to the VNT. In fact, other language measures, such as the NAB Naming Test, as well as other well-established neuropsychological assessment measures, such as the CVLT-II, also showed significant correlations with almost all tests administered. For example, CVLT-II Trials 1–5 Total Recall correlated significantly with WAIS-IV Block Design $r_s = .468$ ($p < .001$) and Visual Puzzles $r_s = .513$ ($p < .001$). The reason for these strong correlations is unclear. Given the widespread nature of this phenomenon, future work is needed to further investigate the degree to which the VNT and other neuropsychological measures demonstrate a clear pattern of convergent and discriminant validity. This could help clarify whether the VNT is a specific measure of naming, as hypothesized, or if it could also be used to detect general cognitive dysfunction.

The present study additionally sought to develop a discontinue rule for the VNT. Rather than attempting to find a number of consecutive missed items to establish a discontinue rule (e.g., 5 in a row), this study sought to determine an absolute number of nonconsecutive items missed at which point examiners can stop administering the test, in order to avoid excess frustration for patients. Here, it was determined that a score of 34.5 can be reliably used to distinguish between major NCD and mild NCD with 95% specificity. This score also has strong positive predictive power, at 88%. Further discrimination of scores below 34.5 (for example, comparing a score of

29 to a score of 30) appears to add little diagnostic utility. Therefore, it is proposed that once a participant has missed 16 items, the test can be discontinued in order to minimize evaluation time and patient distress, while maximizing diagnostic decision accuracy.

The present study also sought to examine whether VNT performance differed based on demographic variables. In the creation of the VNT, care was taken to select items that were relatively free of relationships with education and were as culture-free as possible. However, in the present sample, VNT performance was significantly related to education and race. A similar pattern was observed for the NAB Naming Test, suggesting that this limitation is not specific to the VNT. Nevertheless, this should be considered when interpreting data in a clinical setting. At this time, regression-based norms for the VNT account for variation by education (Wynn et al., 2020). However, normative data are not yet available that account for racial differences, which could be mediated by linguistic and cultural differences (i.e. lack of exposure/familiarity to stimuli). Future work should determine whether VNT items show bias related to particular racial or bilingual/multilingual groups to ensure the measure's clinical utility.

Limitations

In the present study, patients were separated into groups based on the presence of a neurocognitive disorder. However, each group included patients with various etiologies. While impaired naming is characteristic of some causes of neurocognitive disorder, such as Alzheimer's disease, naming may be spared in other conditions, such as Lewy body disease (Braaten et al., 2006; Salmon & Bondi, 2009). While the present work suggests that the VNT can be used to detect NCDs more broadly, it is possible that the sensitivity and specificity of the VNT may be found to be higher in detecting specific etiologies of NCD, such as Alzheimer's disease. This could be an area of future work, and exploration with autopsy-defined etiologies would be optimal.

The limited diversity of the participants in the present study is a significant limitation. Our sample contains an uneven distribution of gender, with significantly more males than females. While preliminary evidence in our study as well as Wynn et al. (2020) does not indicate a difference in VNT performance by gender, additional data would be beneficial. Additionally, while the current study combined data from two geographic regions (Midwest United States and Southern United States), there was still limited racial diversity. Specifically, while Black participants made up 21% of the present study, other racial minorities (Hispanic, Asian American, and Native American) were underrepresented and made up only 3% of the sample. Future work should continue to investigate the use the VNT in a more demographically diverse sample. Future work should also include the development of a similar naming test in other languages such

as Spanish and address the possible impact of bi/multi-lingualism on performance.

An additional limitation is the variation in sample size for our exploratory analysis of convergent and divergent validity. As the present study used clinical data, test batteries varied across participants. Therefore, analyses of convergent and divergent validity typically did not include the entire sample of participants, potentially limiting power for these analyses. Future work investigating convergent and divergent validity with larger clinical samples may be beneficial.

CONCLUSION

Word finding is a common complaint among older adults. Impaired confrontation naming is included in the diagnostic guidelines for the detection of major NCD (American Psychiatric Association, 2013; McKhann et al., 2011) and is considered a key part of the neuropsychological assessment battery. While naming is often assessed using picture-naming tasks, the present study provides compelling support for the VNT, an auditory-based measure, in clinical settings. This measure can be particularly useful in tele-neuropsychological evaluations, when the patient is unable to travel to a clinic or when the sharing of test materials must be minimized, such as during the COVID-19 pandemic.

FINANCIAL SUPPORT

None.

CONFLICTS OF INTEREST

None.

REFERENCES

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.).
- Belleville, S., Fouquet, C., Hudon, C., Zomahoun, H.T.V., & Croteau, J. (2017). Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: A systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 328–353. <https://doi.org/10.1007/s11065-017-9361-5>
- Benton, A.L. & Hamsher, K. de S. (1976). *Multilingual Aphasia Examination* (Revised). Iowa City, IA: University of Iowa.
- Blumenfeld, H. (2010). *Neuroanatomy Through Clinical Cases* (2nd ed.). Sunderland, MA: Sinauer Associates.
- Braaten, A.J., Parsons, T.D., McCue, R., Sellers, A., & Burns, W.J. (2006). Neurocognitive differential diagnosis of dementing diseases: Alzheimer's Dementia, Vascular Dementia, Frontotemporal Dementia, and Major Depressive Disorder. *International Journal of Neuroscience*, 116(11), 1271–1293. <https://doi.org/10.1080/00207450600920928>
- Brearly, T.W., Shura, R.D., Martindale, S.L., Lazowski, R.A., Luxton, D.D., Shenal, B.V., & Rowland, J.A. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology*

- Review, 27(2), 174–186. <https://doi.org/10.1007/s11065-017-9349-1>
- Chan, T., Friedman, D.S., Bradley, C., & Massof, R. (2018). Estimates of incidence and prevalence of visual impairment, low vision, and blindness in the United States. *JAMA Ophthalmology*, 136(1), 12–19. <https://doi.org/10.1001/jamaophthalmol.2017.4655>
- Delis, D., Kaplan, E., & Kramer, J. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: The Psychological Corporation.
- Delis, D., Kramer, J., Kaplan, E., & Ober, B. (2000). *California Verbal Learning Test* (2nd ed.). Manual. San Antonio, TX: The Psychological Corporation.
- Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The Assessment of Aphasia and Related Disorders* (3rd ed.). Austin, TX: Pro-Ed.
- Katsumata, Y., Mathews, M., Abner, E.L., Jicha, G.A., Caban-Holt, A., Smith, C.D., . . . Fardo, D.W. (2015). Assessing the discriminant ability, reliability, and comparability of multiple short forms of the Boston naming test in an Alzheimer's disease center cohort. *Dementia and Geriatric Cognitive Disorders*, 39, 215–227. <https://doi.org/10.1159/000370108>
- Lezak, M.D., Howieson, D.B., Bigler, E.D., & Tranel, D. (2012). *Neuropsychological Assessment* (5th ed.). New York, NY: Oxford University Press.
- McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., . . . Phelps, C.H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's and Dementia*, 7(3), 263–269. <https://doi.org/10.1016/j.jalz.2011.03.005>
- Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)*. San Antonio, TX: Psychological Corporation.
- Randolph, C. (2012). *RBANS Update Manual*. San Antonio, TX: The Psychological Corporation.
- Salmon, D.P. & Bondi, M.W. (2009). Neuropsychological assessment of dementia. *Annual Review of Psychology*, 60, 257–282. <https://doi.org/10.1146/annurev.psych.57.102904.190024.Neuropsychological>
- Spreeen, O. & Benton, A.L. (1969). *Neurosensory Center Comprehensive Examination for Aphasia*. Victoria, British Columbia: University of Victoria Neuropsychology Laboratory.
- Stasencko, A., Jacobs, D.M., Salmon, D.P., & Gollan, T.H. (2019). The Multilingual Naming Test (MINT) as a measure of picture naming ability in Alzheimer's Disease. *Journal of the International Neuropsychological Society*, 25(8), 821–833. <https://doi.org/10.1017/S1355617719000560>
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245.
- Stern, R.A. & White, T. (2003). *Neuropsychological Assessment Batter: Administration, Scoring, and Interpretation Manual*. Lutz, FL: Psychological Assessment Resources.
- Wechsler, D. (2008). *Manual for the Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: The Psychological Corporation.
- Weissberger, G.H., Strong, J.V., Stefanidis, K.B., Summers, M.J., Bondi, M.W., & Stricker, N.H. (2017). Diagnostic accuracy of memory measures in Alzheimer's dementia and mild cognitive impairment: A systematic review and meta-analysis. *Neuropsychology Review*, 27(4), 354–388. <https://doi.org/10.1007/s11065-017-9360-6>
- Wynn, M.J., Sha, A.Z., Lamb, K., Carpenter, B.D., & Yochim, B.P. (2020). Performance on the Verbal Naming Test among healthy, community-dwelling older adults. *Clinical Neuropsychologist*, 34(5), 956–968. <https://doi.org/10.1080/13854046.2019.1683232>
- Yochim, B.P., Beaudreau, S.A., Kaci Fairchild, J., Yutsis, M.V., Raymond, N., Friedman, L., & Yesavage, J. (2015). Verbal naming test for use with older adults: Development and initial validation. *Journal of the International Neuropsychological Society*, 21(3), 239–248. <https://doi.org/10.1017/S1355617715000120>
- Yochim, B.P., Rashid, K., Raymond, N.C., & Beaudreau, S.A. (2013). How frequently are words used on naming tests used in spoken conversation? *Clinical Neuropsychologist*, 27(6), 973–987. <https://doi.org/10.1080/13854046.2013.797501>