Published in final edited form as:

Clin Neuropsychol. 2015 October ; 29(7): 1034–1052. doi:10.1080/13854046.2015.1122085.

Reading Ability as an Estimator of Premorbid Intelligence: Does It Remain Stable Among Ethnically Diverse HIV+ Adults?

J. Pat Olsen^{1,3}, Robert P. Fellows¹, Monica Rivera-Mindt^{1,2,3}, Susan Morgello^{1,4}, and Desiree A. Byrd^{1,2} for the Manhattan HIV Brain Bank

¹Department of Neurology, The Icahn School of Medicine at Mount Sinai, New York, NY

²Department of Psychiatry, The Icahn School of Medicine at Mount Sinai, New York, NY

³Department of Psychology, Fordham University, New York, NY

⁴Department of Neuroscience, The Icahn School of Medicine at Mount Sinai, New York, NY

Abstract

The Wide Range Achievement Test, 3rd edition, Reading-Recognition subtest (WRAT-3 RR) is an established measure of premorbid ability. Furthermore, its long-term reliability is not well documented, particularly in diverse populations with CNS-relevant disease. Objective: We examined test-retest reliability of the WRAT-3 RR over time in an HIV+ sample of predominantly racial/ethnic minority adults. Method: Participants ($N = 88$) completed a comprehensive neuropsychological battery, including the WRAT-3 RR, on at least two separate study visits. Intraclass correlation coefficients (ICCs) were computed using scores from baseline and follow-up assessments to determine the test-retest reliability of the WRAT-3 RR across racial/ethnic groups and changes in medical (immunological) and clinical (neurocognitive) factors. Additionally, Fisher's Z tests were used to determine the significance of the differences between ICCs. Results: The average test-retest interval was 58.7 months ($SD=36.4$). The overall WRAT-3 RR test-retest reliability was high ($r = .97, p < .001$), and remained robust across all demographic, medical, and clinical variables (all r 's $> .92$). Intraclass correlation coefficients did not differ significantly between the subgroups tested (all Fisher's $Z p$'s $> .05$). Conclusions: Overall, this study supports the appropriateness of word-reading tests, such as the WRAT-3 RR, for use as stable premorbid IQ estimates among ethnically diverse groups. Moreover, this study supports the reliability of this measure in the context of change in health and neurocognitive status, and in lengthy inter-test intervals. These findings offer strong rationale for reading as a "hold" test, even in the presence of a chronic, variable disease such as HIV.

Keywords

reliability; reading tests; premorbid IQ; HIV/AIDS; low literacy

Correspondence concerning this article should be addressed to James P. Olsen, Fordham University, 441 East Fordham Road, New York, NY 10458. jolsen10@fordham.edu. Phone: 631-525-9335. Fax: 212-876-9547..

The authors report no conflicts of interest.

Introduction

The estimation of premorbid intellectual functioning is an integral step in the neuropsychological assessment and diagnostic process. While there are various approaches to estimating premorbid functioning, the “hold” approach, using reading level, is one of the most widely used in clinical and research settings due to the tests’ ease of administration, patient tolerability, and low costs (Barona, Reynolds, & Chastain, 1984; Lezak, Howieson, & Loring, 2004; Schoenberg, Scott, Duff, & Adams, 2002). Tests assessing reading level, which measure the ability to read phonetically irregular words, rely more heavily on previous knowledge than current cognitive functioning (Franzen, Burgess, & Smith-Seemiller, 1997), are moderately to strongly correlated with intelligence (Griffin, Mindt, Rankin, Ritchie, & Scott, 2002), are considered temporally stable (Casaletto et al., 2014; Crawford, Parker, Stewart, Besson, & De Lacey, 1989; Harvey et al., 2006; Morrison, Sharkey, Allardyce, Kelly, & McCreddie, 2000; Uttl, 2002), and make use of abilities that are typically only mildly affected by many forms and degrees of cerebral injury or pathology (Casaletto et al., 2014; Green, Melo, Christensen, Ngo, Monette, & Bradbury, 2008; Maddrey, Cullum, Weiner, & Filley, 1996; Patterson, Graham, & Hodges, 1994; Stebbins, Wilson, Gilley, Bernard, & Fox 1990; Willshire, Kinsella, & Prior, 1991).

The clinical utility of a word-reading test to estimate premorbid intelligence is contingent upon its uniform performance as a measure of IQ, as well as its stability over time, in the groups for which it is utilized. However, it has been documented that reading scores in racial/ethnic minority populations fall below those of non-Hispanic White populations matched for years of education at each grade level (Cosentino, Manly & Mungas, 2007; Boekamp, Strauss & Adams, 1995). Specifically, this research highlights discrepant reading scores across non-Hispanic White and African American persons (Boekamp, Strauss & Adams, 1995), and across non-Hispanic White, African American and Hispanic/Latino adults (Cosentino, Manly & Mungas, 2007). Additionally, research suggests that in populations with low quality of education, reading tests will underestimate premorbid function (Ryan et al., 2008). For example, in a largely minority, HIV+ population, a non-verbal test of intellectual function, the General Ability Measure for Adults (GAMA), was a better predictor of neuropsychological functioning than reading scores for low literacy individuals (Ryan et al., 2008). Regarding group differences in test-retest reliability, research examining the stability of reading scores over a 1-year interval found that African American participants showed a trend toward greater change scores compared to non-Hispanic White participants (Orme, Johnstone, Hanks, & Novack, 2004). This trend approached significance, suggesting that race may affect the test-retest reliability of word reading tests. It remains unknown if stability of reading scores varies depending on race or ethnicity.

There are several reading tasks that have been used to estimate premorbid cognitive ability, including the National Adult Reading Test (NART; Nelson & McKenna, 1975), the North American Adult Reading Test (NAART; Spreen & Strauss, 1991), the Wide Range Achievement Test-Reading Recognition subtest (WRAT-RR, Jastak & Wilkinson, 1984; Wilkinson, 1993; Wilkinson & Robertson, 2006) and the Wechsler Test of Adult Reading (WTAR, Wechsler, 2001). Although the validity of these measures as achievement tests has

been generally well established (Crawford, Deary, Starr & Whalley, 2001; Uttl, 2002; Whitney, Shepard, Mariner, Mossbarger, & Herman, 2010), as discussed above, reading tests may be poor predictors of IQ, particularly at high and low extremes (Griffin et al., 2002; Ryan et al., 2008; Sprinks et al., 2009).

Test Stability in Racial/Ethnic Minority Groups

Research examining WRAT reading test performance across a number of clinical adult samples indicates relatively stable performance among psychiatric inpatients, individuals with substance use disorders, elders suffering from neurocognitive impairment, and individuals recovering from traumatic brain injuries (Ashendorf, Jefferson, Green, & Stern, 2009; Dura, Myers, & Freathy, 1989; Harvey et al., 2006; Johnson, Fisher, Rhodes, & Booth, 1996; Johnstone & Wilhelm, 1996; McCaffrey, Duff, & Westervelt, 2000; Orme, Johnstone, Hanks, & Novack, 2004; Woodward, Santa-Barbara, & Roberts, 1975). However, it is problematic that data supporting the long-term test-retest reliability of word recognition reading tests among ethnically diverse individuals and adults with low reading levels are scarce (Harvey et al., 2006; Morrison et al., 2000). Most prior studies of reading test reliability have examined racially/ethnically homogeneous samples consisting largely of non-Hispanic White participants with at least a high school diploma who read at or above a high school level (Casaletto et al., 2014; Johnson, Fisher, Rhodes, & Booth, 1996; Johnstone & Wilhelm, 1996).

To date, no test-retest reliability studies of WRAT reading ability have examined predominantly racial/ethnic minority samples whose reading level is below average, leaving the stability of these measures among these groups in question. Empirical support for the psychometric soundness of this measure is particularly important given the compelling evidence to suggest that the WRAT is a better method of measuring premorbid intelligence than years of education in ethnic minorities and persons with low socioeconomic status (Cosentino, Manly, & Mungas, 2007; Del Ser, Gonzalez-Montalvo, Martinez-Espinosa, Delgado-Villaplos, & Bermejo, 1997; Manly, Jacobs, Touradji, Small, & Stern, 2002; Ryan, Baird, Mindt, Byrd, Monzones, & Morgello, 2005). Moreover, the WRAT is among the most routinely used measures by clinicians and researchers to estimate premorbid functioning and deserves empirical validation of its psychometric properties (Smith-Seemiller, Franzen, Burgess, & Prieto, 1997). Thus, there is need for evidence supporting the test-retest reliability of the WRAT among under-studied populations (i.e., racial/ethnic minorities, persons with low reading levels); especially given the evidence of discrepant reading scores between non-Hispanic White participants and racial/ethnic minority participants (Cosentino, Manly & Mungas, 2007; Boekamp, Strauss & Adams, 1995).

WRAT Stability in Diseased Samples

Longitudinal research examining the test-retest reliability of reading ability indicates that WRAT reading test performance remains relatively stable across a number of clinical samples and across distinct injury/disease severity groups within these samples, including mild, moderate and severe traumatic brain injury groups (Orme, Johnstone, Hanks, & Novack, 2004) and control, mild cognitive impairment and possible/probable Alzheimer's disease groups (Ashendorf, Jefferson, Green, & Stern, 2009). However, there is also

evidence to suggest that word reading ability, as measured via WRAT performance, declines with Huntington's Disease (HD) progression (O'Rourke et al., 2011). The unique report of a significant change in WRAT performance only for Huntington's Disease versus other dementia is important in the context of diseases such as HIV, given that both diseases involve frontostriatal circuitry. Moreover, unlike neurodegenerative disorders such as Alzheimer's and Huntington's disease, the course of HIV-associated neurocognitive disorders (HAND) is not predictable. Individuals with HAND may demonstrate significant recovery, worsening, remain stable or fluctuate in their cognitive and medical status (Heaton et al., 2015). Although word-reading tests like the WRAT are frequently utilized as premorbid measures in individuals with HIV (Everall et al., 2009; Heaton et al., 2010; Woods et al., 2004), there is very little evidence establishing the test-retest reliability of this measure in this population (Casaletto et al., 2014). Therefore, longitudinal research examining the reliability of reading ability is required in populations characterized by fluctuating medical and/or clinical progression (i.e., HIV), in order to determine the applicability of these measures as true "hold" tests that are resistant to the cognitive sequelae of disease and thus appropriate for use as a premorbid indicator late in disease stage.

Recently, the test-retest reliability of the WRAT was examined in a sample of individuals with HIV (Casaletto et al., 2014). In this study, a comprehensive neuropsychological battery, including the WRAT-4, was administered at two visits (M interval = 14.4 months) to HIV+ participants. The results showed that test-retest differences in reading performance were minor, despite improved disease and neurocognitive functioning. While this was the first study to show evidence of test-retest reliability of the WRAT reading test in this particular population, there are several limitations that should be considered, including a relatively small ($n = 48$) and highly homogenous sample (92% men and 67% non-Hispanic White participants) who remained fairly stable or improved in terms of HIV disease severity and neurocognitive functioning. Moreover, this sample's WRAT reading level fell within the average range, and it is unclear whether those with low reading levels would demonstrate similarly robust reliability.

While the aforementioned study provided valuable support for the WRAT as a stable premorbid indicator (Casaletto et al., 2014), replication of these findings among larger, more demographically representative (i.e., racially/ethnically diverse) HIV+ samples that are exhibiting greater fluctuations in disease outcomes is necessary. Further, given the chronic nature of HIV disease progression, extended time intervals (i.e., multiple years) are particularly germane for establishing the WRAT as an adequate "hold" test throughout the course of disease.

Longer Intervals for Chronic Conditions

No prior published studies examining the stability of word-reading tests have assessed individuals in a follow-up assessment more than 7.5 years from baseline, leaving the longer-term reliability of these measures unknown. Specifically, the WRAT-3 testing manual's (Wilkinson, 1993) report of strong test-retest reliability ($r = .98$) is based on a 37-day interval. Most replication studies strengthened this manual's report of strong reading test stability by extending the scope of analysis to approximately 6 to 28 months (Ashendorf et

al., 2009; Casaletto et al., 2014; Johnstone & Wilhelm, 1996; Smith, Roberts, Brewer, & Pantelis, 1998). The longest published test-retest interval in a study of reading test stability is 7.5 years, where stability was explored in a sample of individuals with schizophrenia (Morrison et al., 2000). However, it is unclear whether word-reading tests, such as the WRAT, possess adequate stability (i.e., test-retest reliability) within the context of longer time intervals (over 7.5 years) in a neurologically at-risk HIV+ sample.

The issue of longer time intervals is especially salient given that the epidemiology of HIV has changed significantly over the years. Individuals with HIV are now living significantly longer due to major advances in treatment (Chambers et al., 2014). As a result, now that HIV is more of a chronic disease, individuals display more fluctuation in their disease characteristics over time, rather than progressive decline (Woods, Moore, Weber & Grant, 2009). This disease fluctuation over extended time becomes more concerning given that word-reading ability appear to decline with HD progression (O'Rourke et al., 2011). Thus, it is particularly important to assess whether similar results may be found among individuals who are HIV+ and are exhibiting variable disease progression over the course of their disease.

Study Aims

To address gaps in the extant literature, the current study aimed to examine the test-retest reliability of WRAT-3 reading test based on racial/ethnic minority status, the presence of disease fluctuations in medical (immunological) status and neurocognitive functioning, and the length of interval between assessments.

Methods

Participants

Eighty-eight study participants were drawn from the Manhattan HIV Brain Bank (MHBB; U01MH083501), a longitudinal observational organ donation study that includes annual neurologic, neurocognitive, and psychiatric examinations of HIV+ participants who have given consent for post-mortem organ donation for research purposes. MHBB participation inclusion criteria include: (1) advanced HIV disease, (2) a CD4 count less than or equal to 50 cells/mm³ for at least a 3-month period of time, (3) substantive risk for imminent mortality in the judgment of the participant's primary physician, including another disease without effective therapy (e.g., having a progressive multifocal leukoencephalopathy, lymphoma, disseminated *Mycobacterium avium*–intracellulare, wasting, AIDS dementia complex, cytomegalovirus end-organ disease, visceral Kaposi sarcoma, congestive heart failure, or a serum albumin level of less than 3.2 g/dL.). All participants were English-speaking. Hispanic/Latino participants all reported being English dominant and were mainly of Puerto Rican or Dominican descent. For further MHBB inclusion criteria see Ryan et al., 2005.

For this study analysis, MHBB participants were excluded if they were known to have a condition (i.e., CMV retinopathy, blindness) associated with impaired visual acuity that precluded accurate assessment of reading ability. Inclusion criterion for this study included

the completion of the neuropsychological test battery and WRAT-3 RR test at baseline and again after a minimum of 12 months. Data reported in this manuscript were obtained in compliance with the Mount Sinai School of Medicine Institutional Review Board. Additional demographic and medical characterization of the sample is provided in the Results section.

Procedure

Reading level—The Reading Recognition subtest of the Wide Range Achievement Test-Version 3 (Wilkinson, 1993) was administered to assess reading level as part of a comprehensive battery of neuropsychological tests (detailed below). Per standard administration practices, participants were asked to pronounce words printed on a card and if they were unable to correctly pronounce ten consecutive words, to name letters also printed on the card. At the first improperly pronounced word, the participant was queried to repeat that word. No subsequent feedback was provided. Words are listed in order of decreasing familiarity and increasing phonological complexity. Raw scores were converted to standard scores (SS) or grade level equivalent, according to age-adjusted formulas in the testing manual, for use in analyses.

Educational level—As part of the neuropsychological evaluation, participants were asked to report their highest level of educational achievement. Total years of education were the sum of years completed in elementary, high school and college. Those who did not attain a high-school diploma, but obtained a GED (General Education Development) certificate were assigned the number of actual years of completed formal schooling unless they had advanced to complete some college, wherein they would have been credited for each year of college completed.

Neuropsychological assessment—Participants were administered a neuropsychological test battery, which assessed a broad range of cognitive abilities sensitive to HIV impairment (Woods et al., 2004). All assessments were administered by trained psychometrists, supervised by a clinical neuropsychologist. Table 1 summarizes the two-hour battery, which consists of measures assessing the following seven theoretically-derived ability domains: motor skills, processing speed, abstraction/executive functioning, learning, memory, attention/working memory, and verbal fluency (Woods et al., 2004). Table 1 also presents the norms used to convert raw scores to age-, education-, gender- and where available, ethnicity-corrected *T*-scores using published procedures based upon large normative data sets. Domain scores were derived from the mean *T*-scores of the individual tests in that particular domain, and the global domain score is the mean of all of the individual neuropsychological test *T*-scores.

Clinical ratings of neurocognitive impairment were assigned by a clinical neuropsychologist to each domain using a scale that ranges from one (above average) to nine (severely impaired; Heaton, et al., 1994; Woods et al., 2004). Global clinical ratings were assigned based upon the ratings of the individual domains. A clinical rating of 5 was used as a cutoff to identify cognitive impairment, consistent with established guidelines for HIV-associated NP impairment (Antinori et al., 2007; Grant & Atkinson, 1995).

In routine biannual MHBB study visits, all participants complete the core NP battery with alternate forms of tests, when available. However, the same form of the WRAT-3 RR was only administered at the baseline visit (Time 1) and during a single assessment at or after the 12-month study visit. Thus, while all participants were administered WRAT-3 at baseline, the point at which they received the second administration (Time 2) of the test varied considerably. Nonetheless, each participant received only two exposures to this test.

Laboratory values—Current absolute CD4 count and viral load were obtained either through performance of on-study laboratories or medical record abstraction. Urine toxicology screened for amphetamine, barbiturates, benzodiazepines, cannabinoids, cocaine, opiates, phencyclidine, methadone, and propoxyphene; the illicit status of substances was assessed by a physician upon review of prescribed medications.

Subgroup classifications—Reliability of the WRAT-3 RR was measured for the sample overall, and separately for subgroups based on changes in CD4 count and viral load, illicit substance toxicology, and change in cognitive status, as well as general demographics (i.e., length of test-retest interval, race/ethnicity and years of education). Three test-retest interval subgroups were classified by length of time between assessments: 1 to 3 years, 3 to 6 years and 6+ years. Education subgroups were created by dividing participants into the following categories: “some high school or less,” (<12 years), “high school,” (12 years), and “post high school” (>12 years). Table 2 outlines the change parameters from baseline to follow-up for a number of clinical variables.

Data analysis—Descriptive statistics were calculated on all variables (i.e., race/ethnicity, education, inter-test interval, CD4 count, viral load, illicit substance toxicology, and neurocognitive ratings). Chi square analyses were performed to compare the proportion of subgroup classifications relative to other subgroups. The WRAT-3 RR raw scores from both visits were converted to standard scores using normative data correcting for age (Wilkinson, 1993). Test-retest reliability over the intervals was calculated by means of intraclass correlation (ICC), (Blalock, 1979; Cicchetti, 1994; Fleiss, 1981). The ICC model “2-way mixed” type, “consistency,” was used; ICC analyses also yielded an unbiased estimate of reliability, which reflects the consistency of the baseline (Time 1) assessments (Fleishman & Benson, 1987). Additionally, paired *t*-tests were performed for each subgroup to ensure that mean scores did not change significantly between Time 1 and Time 2 assessments.

In order to identify which participants remained stable, improved or declined neurocognitively between Time 1 and Time 2, a standard regression-based change score method (sRCS) was utilized (Cysique et al., 2011). This approach, which corrects for repeated measures (e.g., practice effects, test reliability, regression to the mean), uses a previously identified normative sample that includes both HIV-seronegative adults and neuromedically stable adults with HIV. Individual z-scores at different times were computed by dividing the difference between predicted and obtained follow-up scaled scores by the error term of the regression model. The resulting z-score reflects how well or poorly the participant did at follow-up, relative to expectations made from his/her baseline (or previous) score and other variable specific baseline predictors. The difference between the predicted retest performance and actual retest performance on each of the individual

neuropsychological tests was then converted into standard, regression-based z scores. Standard change scores (sRCS) for global test performance were created by averaging these z scores. Lastly, confidence intervals of 90% were utilized to determine significant neurocognitive change with the top 5% considered significant improvement and the bottom 5% considered significant decline. The clinical purpose of these intervals is to establish if the discrepancies between a predicted score established at baseline (i.e., time 1) and the obtained scores at follow-up (i.e., time 2) are beyond what would be expected. These conservative confidence intervals were utilized to remain consistent with prior studies utilizing this method (Cysique et al., 2011; Crawford & Garthwaite, 2006; Casaletto et al., 2014; Temkin, Heaton, Grant & Dikmen, 1999). Thus, to summarize, this method was used to identify subgroups of cognitively stable participants, and participants who improved neurocognitively and declined neurocognitively. For more details on this statistical approach and normative data, please see Cysique et al. (2011).

Results

Demographic and medical characteristics

Table 3 summarizes participant characteristics. Of the 88 participants in this study, most were of racial/ethnic minority background: 53.4% African American and 31.8% Hispanic/Latino. The average age at Time 1 was 46.9 years ($SD=6.8$), while mean years of education was 11.7 ($SD=2.6$), with a range of 6 to 18 years. In this sample, 50% completed less than 12 years of school, 21.6% completed high school and 28.4% received education post-high school. Notably, 38.6% of participants had less than an 8th-grade reading level equivalent at Time 1. In terms of neurocognitive functioning, the participants' average level of global and domain specific neurocognitive functioning is summarized in Table 3. In addition, Table 3 illustrates that on average, the sample's global neurocognitive functioning fell within the impaired range (based on Global NP Clinical Ratings), with an overall average rating of 5.8 at Time 1 and 6.1 at Time 2.

While the average test-retest interval was 4.9 years ($SD=3.0$; Range = 1-12.5 years), 36.3% completed the WRAT-3 RR retest 1-3 years after Time 1, 30.7% were tested 3-6 years following, and 33% completed the second administration more than 6 years following Time 1; the longest interval was 12 years. Regarding immunologic status (CD4 count and detectable viral load status), the group's overall profile was similar at both assessments. At Time 1, the median CD4 count was 313.5 and 60.2% of participants had detectable viral loads. At Time 2, the median CD4 count was 364 and 71.6% of participants had detectable viral loads; these differences did not reach statistical significance (all p 's $>.10$). Chi square analyses did not reveal any significant confounds between subgroup classifications (e.g., participants in the lower education group did not receive shorter or longer delays compared to other education subgroups). Furthermore, paired t -test results revealed no significant differences between Time 1 and Time 2 assessments for the overall sample, as well as individual subgroups (all p 's $>.05$).

Test-retest stability analyses

Table 4 presents intraclass correlation analyses for test-retest analyses across clinical and demographic characteristics. The intraclass correlation for the entire sample was .97 ($p < .001$). For the sample as a whole ($N=88$), the average WRAT-3 RR standard scores increased from Time 1 ($M = 83.9$, $SD=16.6$) to Time 2 ($M = 85.3$, $SD=17.4$) by 4.7 points ($SD = 3.6$), with a range of 0-13 points. Only 13.5% of cases had Time 2 WRAT-3 RR scores that were 10 or more standard score points away from their baseline (Time 1) WRAT-3 RR scores. As predicted, intraclass correlations remained high (all p values $> .92$) when compared across all clinical and demographic characteristics.

When examining potential stability variation across ethnicity, ICCs for participants who are African American, Hispanic, and non-Hispanic White were .96, .97, and .95, respectively (all p 's $< .001$). Analysis of test-retest reliability by education level revealed similar ICC estimates for all three education groups (all p 's $< .001$). In an analysis of the degree to which length of test-retest interval may have impacted test stability, results indicate intraclass correlation coefficients were strong across test intervals. Specifically, ICCs were .98 for the "1 to 3 year" and "3 to 6 year" interval subgroups, and .95 for the "6+ year" interval subgroup (all p 's $< .001$).

When changes in immunologic status were examined, results revealed that test stability was no more variable for persons whose immunologic status improved or declined over the interval. Intraclass correlation coefficients remained high for both CD4 count subgroups (.97 for the "CD4 stable" subgroup and .96 for the "CD4 change" subgroup; both p 's $< .001$), as well as for both viral load subgroups (.98 for "viral load stable" and .95 for "viral load change"; both p 's $< .001$). Similarly, test-retest reliability was consistent across illicit substance toxicology status (.97 for those with consistent toxicology reports and .96 for the "toxicology change" subgroup; both p 's $< .001$).

In terms of neurocognitive functioning, sRCS z -scores evidenced variability across the group ranging from -2.8 to 1.6 ($M = -0.33$, $SD = 0.31$), twenty-three participants exhibited significant decline from test to retest, eight participants exhibited significant improvement and fifty-seven participants remained stable, using the 90% CI criterion. ICCs for cognitive status were .94, .95 and .96 (all $p < .001$) for participants who were cognitively stable, exhibited cognitive improvement, and exhibited cognitive decline, respectively. Therefore, despite the fact that 35% of participants changed in their neurocognitive status, the WRAT-3 RR performance was not associated with global or domain-based neuropsychological changes (i.e., sRCS) across the two visits.

Table 5 summarizes the results of a series of Fisher's Z transformation analyses, which were computed to assess the significance of the difference between intraclass correlation coefficients among the subgroups examined in this study (e.g., medical characteristics, neurocognitive function over time, race/ethnicity, education, and test-retest interval). Importantly, there were no significant differences in correlation coefficients between the subgroups for any of demographic, medical or clinical characteristics that were examined (all p 's $> .05$).

Discussion

The WRAT-3 Reading Recognition subtest is often recommended for use in neuropsychological evaluations as an estimate of premorbid ability and education quality, especially for racial/ethnic minorities for whom years of education may not be reflective of actual premorbid intellectual functioning, secondary to differences in the quality of education received (Manly et al., 2002). In contrast to previous studies examining the test-retest reliability of word-reading measures of premorbid functioning over a relatively shorter time period, the present study is the first to examine the reliability in a sample of predominantly ethnic minority HIV+ participants over an extended test-retest interval of almost five years on average. The results from this study support the strong test-retest reliability of this instrument for this cohort, even after very long intervals. In fact, this is the first study to examine this issue with an HIV+ cohort over such extended time intervals. Furthermore, stability coefficients were high across all clinical and demographic factors analyzed in this study. Thus, the present study extends the prior literature by demonstrating high reliability of the WRAT-3 RR test in a sample of HIV+ individuals with greater disease fluctuation and racial/ethnic diversity, and in a sample that is neurocognitively impaired on average who exhibit relatively lower average reading level scores than participants in prior test-retest reliability studies.

The current study is important because it is the first to examine long-term reading stability in a diverse HIV+ sample consisting primarily of African and Hispanic/Latino participants, whereas previous stability studies have examined predominantly non-Hispanic White and/or HIV-negative samples (Casaletto et al., 2014; Johnstone & Wilhelm, 1996; Orme et al., 2004). In contrast to previous research that has reported racial/ethnic differences in average WRAT change scores (Orme et al., 2004), the results of the current study revealed minimal, non-significant differences in WRAT-3 RR stability correlations between race/ethnic groups. Importantly, these results support the applicability of this word-reading test as an estimate of premorbid functioning within diverse patient populations. Specifically, the current study supports the utilization of the WRAT-3 RR as a “hold” test in African American and Hispanic/Latino individuals. This is particularly relevant given that racial/ethnic minorities are at increased risk for neurodegenerative diseases such as HIV (CDC, 2012).

One major question, only partially answered in the literature, is whether reading measures are reliable over time in persons with CNS disease, like advanced HIV infection. A unique and important contribution provided by this current study that is absent in prior research (Casaletto et al., 2014) is the inclusion of individuals who exhibited immunological and neurocognitive declines (i.e., significant increases in HIV viral load, significant decreases in CD4 count and significant decreases in sRCS) in the sample. The current study provides strong support for the test-retest reliability of this word reading test for individuals who declined either immunologically or neurocognitively. Moreover, the test-retest reliability coefficients for individuals who exhibited such decline were not significantly different from those who remained stable or improved immunologically or neurocognitively. Contrary to research in Huntington’s Disease (O’Rourke et al., 2011), this study in HIV disease revealed stability in WRAT reading over time despite declines in health and neurocognitive

functioning. It is possible that the reported declines in WRAT reading ability within HD progression are associated with greater motor dysfunction compared to HIV. Although both diseases share disturbances in frontostriatal circuitry, the motor-speech dysfunction found in HD is more prominent than what is typically observed in HIV disease (McCabe, Sheard, & Code, 2007; Hartelius, Carlstedt, Ytterberg, Lillvik, & Laakso, 2003). Striatal pathology in HIV may not be of sufficient severity to produce dysarthria, which may be related to the ability to complete oral reading ability. Nonetheless, it remains unclear whether or not the WRAT reading instability found in HD disease progression is due to motor dysfunction rather than the effects of dementia altogether.

A novel aspect of the current study is the exceptionally extended interval over which the assessments in the current study were completed. Previously, stability performance on word-reading measures of premorbid ability has been reported for a maximum of 7.5 years (Morrison et al., 2000). Given the longitudinal nature of the parent MHBB study, the present investigation was able to employ a broad range of test-retest intervals (i.e., from 1 to 12.5 years) and test whether interval length impacted stability estimates. Test-retest reliability is typically affected by time intervals between assessments, with shorter retest intervals leading to higher reliability coefficients and longer intervals leading to lower coefficients. Thus it was expected that coefficients would decrease as the interval increased (Duff, 2012). We found a minimal, non-significant decrease in stability as time between Time 1 and Time 2 increased. These results suggest that regardless of length of interval between assessments, reliability coefficients for WRAT-3 RR scores remain high. Therefore, the current study offers strong rationale for reading as a “hold” test, even in the presence of chronic HIV infection and disease fluctuation.

The present study is not without limitations. This study was limited by small sample sizes, especially in subgroup analyses, due to our selection of a prolonged test-retest interval. Language dominance was also determined informally, rather than through formal dominance testing. Of note, the WRAT-3 was utilized in this investigation because the parent study began prior to the release of the WRAT-4 (Wilkinson & Robertson, 2006). However, given the substantial overlap between the WRAT-3 and WRAT-4 and its shared validation process, the two versions of the test are similar premorbid estimators (Wilkinson & Robertson, 2006). In terms of future directions, since a newer revision of the WRAT has been published, it will be important to determine whether long-term stability will generalize to the WRAT-4.

Additionally, given that reading tests have been suggested to underestimate premorbid functioning in populations with low quality of education (Ryan et al., 2008), further research investigating the long-term stability of other, nonverbal premorbid estimates administered to HIV+ samples is warranted. For example, the GAMA has been suggested to be a more accurate premorbid estimation method for individuals with low literacy and/or verbal learning disabilities (Ryan et al., 2008). While the current study firmly establishes the robust test-retest reliability of the WRAT-3 RR in an advanced HIV+ cohort, it remains to be determined the degree to which the measure is a valid estimate of premorbid functioning. Future prospective studies with pre-infection data would be required to address this important question.

Despite the limitations noted above, several aspects of this study make unique contributions to the existing literature. This study established that the WRAT-3 RR is a stable and reliable tool for the assessment of reading levels in a racially/ethnically diverse sample that, on average, have low literacy levels, even with lengthy test intervals (up to 12.5 years). Although much data exists on the test characteristics of word-reading tasks in general, many previous studies have included somewhat homogeneous and fairly well-educated samples. The current study's sample differed significantly from previous studies. Our data successfully extend the findings of previous research to an under-represented and under-studied population (i.e., predominately African American and educationally disadvantaged adults) with greater disease fluctuation than prior studies.

Acknowledgements

The authors gratefully acknowledge MHBB volunteers and research staff. Research supported by awards U24MH100931 (to SM) from the National Institutes of Health, and UL1TR000067 from the National Center for Advancing Translational Sciences to the Clinical Research Center of the Icahn School of Medicine at Mount Sinai.

This work was supported by the National Institutes of Health under Grant U24MH100931; and the National Center for Advancing Translational Sciences under Grant UL1TR000067.

References

- Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M, Wojna VE. Updated research nosology for HIV-associated neurocognitive disorders. *Neurology*. 2007; 69:1789–1799. doi: 10.1212/01.WNL.0000287431.88658.8b. [PubMed: 17914061]
- Ashendorf L, Jefferson AL, Green RC, Stern RA. Test-retest stability on the WRAT-3 reading subtest in geriatric cognitive evaluations. *Journal of Clinical and Experimental Neuropsychology*. 2009; 31:605–610. doi: 10.1080/13803390802375557. [PubMed: 18821160]
- Barona A, Reynolds CR, Chastain R. A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Counseling and Clinical Psychology*. 1984; 52:885–887. doi: 10.1037/0022-006X.52.5.885.
- Benedict, RH. Brief visuospatial memory test-revised. Psychological Assessment Resources; Odessa, FL: 1997.
- Benedict RH, Schretlen D, Groninger L, Brandt J. Hopkins verbal learning test – revised: Normative data and analysis of inter-form and test–retest reliability. *The Clinical Neuropsychologist*. 1998; 12:43–55. doi: 10.1076/clin.12.1.43.1726.
- Blalock, HM, Jr.. Social statistics. 2nd. McGraw-Hill; New York: 1979.
- Boekamp JR, Strauss ME, Adams N. Estimating premorbid intelligence in African-American and White elderly veterans using the American version of the North American Reading Test. *Journal of Clinical and Experimental Neuropsychology*. 1995; 17:645–653. [PubMed: 8557806]
- Casaletto KB, Cattie J, Franklin DR, Moore DJ, Woods SP, Grant I, Heaton RK, HNRG Group. The Wide Range Achievement Test-4 Reading subtest “holds” in HIV-infected individuals. *Journal of Clinical and Experimental Neuropsychology*. 2014; 10:2–10. Doi: 10.1080/13803395.2014.960370.
- Centers for Disease Control and Prevention (CDC). HIV among African Americans. 2012. Retrieved from CDC Website: <http://www.cdc.gov/hiv/topics/aa/index.htm>
- Chambers LA, Wilson MG, Rueda S, Gogolishvili D, Shi MQ, Rourke SB, The Positive Aging Review Team. Evidence informing the intersection of HIV, aging, and health: A scoping review. *AIDS Behavior*. 2014; 18:661–675. [PubMed: 24185708]
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:284–290.
- Cosentino S, Manly J, Mungas D. Do reading tests measure the same construct in multiethnic and multilingual older persons? *Journal of the International Neuropsychological Society*. 2007; 13:228–236. doi: 10.1017/S1355617707070257. [PubMed: 17286880]

- Crawford JR, Garthwaite PH. Comparing patients' predicted test scores from a regression equation with their obtained scores: a significance test and point estimate of abnormality with accompanying confidence limits. *Neuropsychology*. 2006; 20(3):259–271. [PubMed: 16719619]
- Crawford JR, Deary IJ, Starr J, Whalley LJ. The NART as an index of prior intellectual functioning: a retrospective validity study covering a 66-year interval. *Psychological Medicine*. 2001; 31:451–458. [PubMed: 11305853]
- Crawford JR, Parker DM, Stewart LE, Besson JAO, Delaney G. Prediction of the WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology*. 1989; 28:267–273.
- Cysique LA, Franklin D, Abramson I, Ellis RJ, Letendre S, Collier A, Heaton RK. Normative data and validation of the regression based summary score for assessing meaningful neuropsychological change. *Journal of Clinical and Experimental Neuropsychology*. 2011; 33(5):505–522. doi: 10.1080/13803395.2010.535504. [PubMed: 21391011]
- Del Ser T, Gonzalez-Montalvo JI, Martinez-Espinosa S, Delgado-Villapalos C, Bermejo F. Estimation of premorbid intelligence in spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition*. 1997; 33:343–356. Retrieved from doi:10.1006/brcg.1997.0877. [PubMed: 9126399]
- Diehr MC, Cherner M, Wolfson TJ, Miller SW, Grant I, Heaton RK, HIV Neurobehavioral Research Center Group. The 50 and 100-item short forms of Paced Auditory Serial Addition Task (PASAT): Demographically corrected norms and comparisons with the full PASAT in normal and clinical samples. *Journal of Clinical Experimental Neuropsychology*. 2003; 25:571–585. doi: 10.1076/jcen.25.4.571.13876. [PubMed: 12911108]
- Duff K. Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Archives of Clinical Neuropsychology*. 2012; 27:248–261. doi:10.1093/arclin/acr120. [PubMed: 22382384]
- Dura JR, Myers EG, Freathy DT. Stability of the Wide Range Achievement Test in an adolescent psychiatric inpatient setting. *Educational and Psychological Measurement*. 1989; 49:253–256. doi: 10.1177/0013164489491028.
- Everall I, Vaida F, Khanlou N, Lazzaretto D, Achim C, Letendre S, Masliah E. Cliniconeuropathologic correlates of human immunodeficiency virus in the era of antiretroviral therapy. *Journal of Neurovirology*. 2009; 15:360–370. doi: 10.3109/13550280903131915. [PubMed: 20175693]
- Flleiss, JL. Statistical methods for rates and proportions. 2nd. Wiley; New York: 1981.
- Fleishman J, Benson J. Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurements*. 1987; 47:891–896.
- Franzen MD, Burgess EJ, Smith-Seemiller L. Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology*. 1997; 12:711–738. doi: 10.1016/S0887-6177(97)00046-2. [PubMed: 14590649]
- Gladysjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK. Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*. 1999; 6:147–178. doi:10.1177/107319119900600204. [PubMed: 10335019]
- Grant, I.; Atkinson, JH. Psychiatric aspects of acquired immune deficiency syndrome. In: Kaplan, HI.; Sadock, BJ., editors. *Comprehensive textbook of psychiatry*. Vol. 2. Williams and Wilkins; Baltimore: 1995. p. 1644-1669.
- Green REA, Melo B, Christensen B, Ngo L, Monette G, Bradbury C. Measuring premorbid IQ in traumatic brain injury: An examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology*. 2008; 30(2):163–172. doi: 10.1080/13803390701300524. [PubMed: 18213530]
- Griffin SL, Mindt M, Rankin EJ, Ritchie AJ, Scott JG. Estimating premorbid intelligence: Comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology*. 2002; 17(5):497–507. doi:10.1016/S0887-6177(01)00136-6. [PubMed: 14592003]
- Hartelius L, Carlstedt A, Ytterberg M, Lillvik M, Laakso K. Speech disorders in mild and moderate Huntington's disease: results of dysarthria assessment of 19 individuals. *Journal of Medical Speech-Language Pathology*. 2003; 34:106–112.

- Harvey PD, Friedman JI, Bowie C, Reichenberg A, McGurk SR, Parrella M, Davis KL. Validity and stability of performance-based estimates of premorbid educational functioning in older patients with schizophrenia. *Journal of Clinical and Experimental Neuropsychology*. 2006; 28:178–192. doi: 10.1080/13803390500360349. [PubMed: 16484092]
- Heaton RK, Clifford DB, Franklin DR, Woods SP, Ake C, Vaida F, CHARTER Group. HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy. *Neurology*. 2010; 75(23):2087–2096. [PubMed: 21135382]
- Heaton RK, Franklin DR Jr, Deutsch R, Letendre S, Ellis RJ, Casaletto K, CHARTER Group. Neurocognitive change in the era of HIV combination antiretroviral therapy: the longitudinal CHARTER study. *Clinical Infectious Disease*. 2015; 60(3):473–480.
- Heaton, R.; Grant, I.; Matthews, CG. Comprehensive norms for an Expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications. Psychological Assessment Resources; Odessa, FL: 1991.
- Heaton, R.; Kirson, D.; Velin, R.; Grant, I.; Group, TH. The utility of clinical ratings for detecting early cognitive change in HIV infection. In: Grant, I.; Martin, A., editors. *Neuropsychology of HIV Infection*. Oxford University Press; New York, NY: 1994. p. 188-206.
- Jastak, S.; Wilkinson, G. The Wide Range Achievement Test: Manual of instructions. Jastak Associates; Wilmington, DE: 1984.
- Johnson ME, Fisher DG, Rhodes F, Booth R. Test-retest stability and concurrent validity of two reading tests with a drug-abusing population. *Assessment*. 1996; 3:111–114. doi: 10.1177/107319119600300202.
- Johnstone B, Wilhelm KL. The longitudinal stability of the WRAT-R Reading subtest: Is it an appropriate estimate of premorbid intelligence? *Journal of the International Neuropsychological Society*. 1996; 2:282–285. doi: 10.1017/S1355617700001296. [PubMed: 9375176]
- Kongs, SK.; Thompson, LL.; Iverson, GL.; Heaton, RK. Wisconsin Card Sorting Test-64 Card Computerized Version. Psychological Assessment Resources; Odessa, FL: 2000.
- Lezak, MD.; Howieson, DB.; Loring, DW. Neuropsychological assessment. 4th. Oxford University Press; New York: 2004.
- Maddrey AM, Cullum CM, Weiner MF, Filley CM. Premorbid intelligence estimation and level of dementia in Alzheimer's disease. *Journal of the International Neuropsychological Society*. 1996; 2:551–555. [PubMed: 9375159]
- Manly JJ, Jacobs DM, Touradj P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and white elders. *Journal of the International Neuropsychological Society*. 2002; 8:341–348. doi:10.1017/S1355617702813157. [PubMed: 11939693]
- McCabe P, Sheard C, Code C. Pragmatic skills in people with HIV/AIDS. *Disability and Rehabilitation*. 2007; 29(16):1251–60. [PubMed: 17654000]
- McCaffrey, RJ.; Duff, K.; Westervelt, HJ. Practitioner's guide to evaluating change with neuropsychological assessment instruments. Plenum Press; New York: 2000.
- Moore RD. Epidemiology of HIV infection in the United States: Implications for linkage to care. *Clinical Infectious Diseases*. 2011; 52:208–213.
- Morrison G, Sharkey V, Allardyce J, Kelly RC, McCreadie RG. Nithsdale Schizophrenia Surveys 21: a longitudinal study of National Adult Reading Test stability. *Psychological Medicine*. 2000; 30:717–720. [PubMed: 10883725]
- Nelson HE, McKenna P. The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology*. 1975; 14:259–267. doi:10.1111/j.2044-8260.1975.tb00178.x. [PubMed: 1182406]
- O'Rourke JJF, Adams WH, Duff K, Byars J, Nopoulos P, Paulsen JS, Beglinger LJ. Estimating premorbid functioning in Huntington's disease: The relationships between disease progression and the Wide Range Achievement Test Reading Subtest. *Archives of Clinical Neuropsychology*. 2011; 26(1):59–66. [PubMed: 21147861]
- Orme DR, Johnstone B, Hanks R, Novack T. The WRAT-3 reading subtest as a measure of premorbid intelligence among persons with brain injury. *Rehabilitation Psychology*. 2004; 49:250–253. doi: 10.1037/0090-5550.49.3.250.

- Patterson K, Graham N, Hodges JR. Reading in dementia of the Alzheimer type: A preserved ability? *Neuropsychology*. 1994; 1:395–407.
- Ryan EL, Baird R, Mind MR, Byrd D, Monzones J, Morgello S, for the Manhattan HIV Brain Bank. Neuropsychological impairment in racial/ethnic minorities with HIV infection and low literacy levels: Effects of education and reading level in patient characterization. *Journal of the International Neuropsychological Society*. 2005; 11:889–898. doi: 10.1017/S1355617705051040. [PubMed: 16519268]
- Ryan EL, Byrd D, Mindt MR, Rausch WJ, Morgello S. Understanding the neuropsychological profile of HIV+ participants with low literacy: role of the General Ability Measure for Adults (GAMA). *The Clinical Neuropsychologist*. 2008; 22:1018–1034. doi: 10.1080/13854040701750883. [PubMed: 18609327]
- Schoenberg MR, Scott JG, Duff K, Adams RL. Estimation of WAIS-III intelligence from combined performance and demographic variables: Development of the OPIE-3. *The Clinical Neuropsychologist*. 2002; 16:426–438. doi: 10.1076/clin.16.4.426.13913. [PubMed: 12822052]
- Smith-Seemiller L, Franzen MD, Burgess EJ, Prieto LR. Neuropsychologists' practice patterns in assessing premorbid intelligence. *Archives of Clinical Neuropsychology*. 1997; 12(8):739–744. doi: 10.1016/S0887-6177(97)00047-4. [PubMed: 14590650]
- Smith D, Roberts S, Brewer W, Pantelis C. Test-retest reliability of the National Adult Reading Test (NART) as an estimate of premorbid IQ in patients with schizophrenia. *Cognitive Neuropsychiatry*. 1998; 3(1):71–80.
- Spreen, O.; Strauss, E. Oxford University Press; New York: 1991. A compendium of neuropsychological tests: Administration, norms, and commentary.
- Sprinks R, McKirgan LW, Arndt S, Caspers K, Yucuis R, Pfalzgraf CJ. IQ estimate smackdown: Comparing IQ proxy measures to the WAIS-III. *Journal of the International Neuropsychological Society*. 2009; 15:590–596. [PubMed: 19573277]
- Stebbins GT, Wilson RS, Gilley D, Bernard BA, Fox JH. Use of the National Adult reading Test to estimate premorbid IQ in dementia. *The Clinical Neuropsychologist*. 1990; 4:18–24.
- Temkin NR, Heaton RK, Grant I, Dikmen SS. Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society*. 1999; 5:357–369. [PubMed: 10349298]
- Uttl B. North American Adult Reading Test: age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*. 2002; 24(8):1123–1137. [PubMed: 12650237]
- Wechsler, D. Wechsler Adult Intelligence Scale – Third Edition: Administration and Scoring Manual. Harcourt Brace; San Antonio, TX: 1997.
- Wechsler, D. Wechsler test of adult reading. The Psychological Corporation; San Antonio, TX: 2001.
- Whitney KA, Shepard PH, Mariner J, Mossbarger B, Herman SM. Validity of the Wechsler Test of Adult Reading (WTAR): effort considered in a clinical sample of U.S. military veterans. *Applied Neuropsychology*. 2010; 17:196–204. doi: 10.1080/09084282.2010.499787. [PubMed: 20799111]
- Wilkinson, GS. WRAT-3: Wide Range Achievement Test administration manual. 3rd. Western Psychological Services; Wilmington, DE: 1993.
- Wilkinson, GS.; Robertson, GJ. Wide range achievement test – Fourth edition: Professional manual. Psychological Assessment Resources; Lutz, FL: 2006.
- Willshire D, Kinsella G, Prior M. Estimating WAIS-R IQ from the National Adult Reading Test: A cross-validation. *Journal of Clinical and Experimental Neuropsychology*. 1991; 13(2):204–216. doi: 10.1080/01688639108401038. [PubMed: 1864911]
- Woods SP, Moore DJ, Weber E, Grant I. Cognitive neuropsychology of HIV-associated neurocognitive disorders. *Neuropsychology Review*. 2009; 19(2):152–168. Doi: 10.1007/s11065-009-9102-5. [PubMed: 19462243]
- Woods SP, Rippeth JD, Frol AB, Levy JK, Ryan E, Soukup VM, Heaton RK. Interrater reliability of clinical ratings and neurocognitive diagnoses in HIV. *Journal of Clinical and Experimental Neuropsychology*. 2004; 26:759–778. doi:10.1080/13803390490509565. [PubMed: 15370374]
- Woodward CA, Santa-Barbara J, Roberts R. Test-retest reliability of the Wide Range Achievement Test. *Journal of Clinical Psychology*. 1975; 31:81–84. doi:10.1002/1097-4679. [PubMed: 1112906]

Table 1

Neuropsychological Battery and Normative Data Sources

Neuropsychological Domain/Tests	Normative Data Source
Motor	
Grooved Pegboard – Dominant Hand	Heaton et al. (1991) ^{1,2,3}
Grooved Pegboard – Non-Dominant Hand	Heaton et al. (1991) ^{1,2,3}
Processing Speed	
Trail Making Test, Part A (TMT-A)	Heaton et al. (1991) ^{1,2,3}
WAIS-III Digit Symbol	Wechsler (1997) ¹
WAIS-III Symbol Search	Wechsler (1997) ¹
Executive Functioning	
Trail Making Test, Part B (TMT-B)	Heaton et al. (1991) ^{1,2,3}
Wisconsin Card Sorting Test (WCST) -Perseverative Responses	Kongs et al. (2000) ^{1,2}
Learning	
Brief Visual Memory Test -Total Recall	Benedict et al. (1997) ¹
Hopkins Verbal Learning Test -Total Recall	Benedict et al. (1998) ¹
Memory	
Brief Visual Memory Test- Delayed Recall	Benedict et al. (1997) ¹
Hopkins Verbal Learning Test- Delayed Recall	Benedict et al. (1998) ¹
Working Memory	
WAIS-III Letter Number Sequencing	Wechsler (1997) ¹
Paced Auditory Serial Addition Task	Diehr et al. (2003) ^{1,2,3,4}
Verbal Fluency	
Controlled Oral Word Association Test	Gladajo et al. (1999) ^{1,2,4}
Reading Level	
Wide Range Achievement Test – Reading 3 rd Edition	Wilkinson (1993) ¹

Note. Normative data provides adjustments for the following demographic characteristics, as indicated:

¹ Age;

² Education;

³ Gender;

⁴ Ethnicity.

Table 2

Description Of Change Parameters For Clinical And Medical Variables

Variable	Change measurement	Subgroup classification
CD4 count	Crossed a boundary value of 200 cells/ μ L in either direction	Stable: CD4 count consistently above or below 200 cells/ μ L at both visits Change: CD4 count increased or decreased past boundary value of 200 cells/ μ L between assessments
Viral load	Crossed a boundary value of detectable virus (50 copies/ μ L)	Stable: consistently undetectable or detectable at both visits Change: viral load at follow-up visit different from baseline visit
Illicit substance use	Urine toxicology test for use of illicit substances	Stable: consistently positive or negative at each visit Change: status at follow-up visit different from baseline visit
Cognitive status	sRCS	Stable: sRCS value remained within 90% CI Decline: Bottom 5% sRCS values Improvement: Top 5% sRCS values

Note. sRCS = standard regression-based change score; CI = confidence interval.

Table 3**Participant Characteristics**

Variable	Time 1 <i>M (SD)</i> or %	Time 2 <i>M (SD)</i> or %
Demographic Characteristics		
Age (years)	46.9 (6.8)	51.9 (7.0)
Sex (% female)	53.4%	--
Race/Ethnicity		
African American (%)	53.4%	--
Hispanic or Latino (%)	31.8%	--
Non-Hispanic White (%)	14.8%	--
Education (years)	11.7 (2.6)	--
Medical Characteristics		
Median CD4 (cells/ μ L)	313.5	364
Detectable HIV RNA (%)	60.2%	71.6%
Illicit substance toxicology (% positive)	25.6%	22.4%
Neurocognitive Characteristics		
(Average T-scores)		
Global	39.4 (7.6)	31.2 (8.7)
Motor	32.6 (9.1)	36.9 (10.9)
Processing Speed	42.3 (8.6)	45.7 (10.4)
Executive Functioning	40.7 (10.9)	46.1 (11.2)
Working Memory	43.0 (9.6)	45.2 (10.7)
Learning	35.4 (11.7)	32.9 (12.0)
Memory	36.0 (13.2)	33.4 (14.3)
Verbal Fluency	49.4 (10.8)	54.6 (11.4)
Reading Level (% <8 th grade equivalence)	38.6%	37.5%
Global Clinical Ratings	5.8 (1.9)	6.1 (2.4)

Note. See Methods section for Global Clinical Rating calculations. $N = 88$.

Table 4

Test-Retest Reliability Characteristics for the Entire Sample (N=88)

Subgroup	N	%	Time 1 WRAT SS <i>M (SD)</i>	Time 2 WRAT SS <i>M (SD)</i>	ρ	ICC 95% CI		UER
						Lower	Upper	
Demographic Characteristics								
Race/Ethnicity								
African American	47	53.4%	80.6 (15.7)	82.2 (16.3)	.96*	.93	.98	.96
Hispanic	28	31.8%	81.1 (15.5)	81.6 (16.4)	.97*	.93	.99	.97
Non-Hispanic White	13	14.8%	101.9 (9.4)	104.4 (10.4)	.95*	.83	.98	.95
Years of education								
Some high school or less (<12 yrs)	44	50.0%	74.6 (14.9)	76.1 (15.2)	.96*	.92	.98	.96
High school (12 yrs)	19	21.6%	89.0 (9.8)	89.4 (10.6)	.93*	.82	.97	.94
Post high school (>12 yrs)	25	28.4%	96.3 (13.6)	98.4 (15.7)	.96*	.91	.98	.96
Test-Retest interval groups								
1-3 yrs	32	36.3%	84.5 (15.1)	84.8 (15.9)	.98*	.96	.99	.98
3-6 yrs	27	30.7%	85.3 (19.4)	87.8 (20.2)	.98*	.96	.99	.98
6+ yrs	29	33.0%	81.9 (15.5)	83.5 (16.5)	.95*	.89	.98	.95
Medical Characteristics								
CD4 change								
CD4 stable	67	76.1%	80.8 (16.0)	82.2 (16.9)	.97*	.95	.98	.97
CD4 change	21	23.9%	93.6 (14.8)	95.3 (15.3)	.96*	.90	.98	.96
HIV RNA viral load								
Viral load stable	58	65.9%	83.7 (18.0)	85.2 (18.9)	.98*	.96	.99	.98
Viral load change	30	34.1%	84.1 (13.6)	85.5 (14.3)	.95*	.90	.98	.96
Illicit substance tox (<i>n</i> = 86)								
Tox stable	63	73.3%	84.6 (17.2)	85.8 (18.1)	.97*	.96	.98	.97
Tox change	23	26.7%	81.4 (15.6)	83.4 (16.3)	.96*	.91	.98	.96
Neurocognitive Characteristics								
sRCS scores								
Cognitively stable	57	64.8%	83.9 (16.4)	85.5 (18.1)	.94*	.89	.97	.95
Cognitive improvement	8	9.1%	91.8 (14.0)	90.5 (17.1)	.95*	.92	.98	.96
Cognitive decline	23	26.1%	79.9 (17.5)	81.4 (17.9)	.96*	.90	.98	.96
Total	88	100%	83.9 (16.6)	85.3 (17.4)	.97*	.96	.98	.97

Note. Time 1 WRAT SS M = Average WRAT-3 RR standard score at first visit; Time 2 WRAT SS M = Average WRAT-3 RR standard score at second visit; ρ = intraclass correlation; ICC 95% CI = confidence interval; UER = unbiased estimate of reliability; sRCS = standard regression-based change score.

* $p < .001$.

Table 5

Fisher's Z Correlation Coefficient Comparisons

Subgroup comparisons	z- value	SEM	p
Race/Ethnicity			
African American vs. Hispanic	-0.60	0.25	0.55
African American vs. Non-Hispanic White	+0.18	0.35	0.86
Hispanic vs. Non-Hispanic White	+0.56	0.37	0.58
Years of education			
Some high school vs. high school	+0.87	0.29	0.38
Some high school vs. post high school	-0.26	0.26	0.79
High school vs. post high school	-0.99	0.33	0.32
Test-Retest interval groups			
1-3 yrs vs. 3-6 yrs	+0.00	0.28	1.00
1-3 yrs vs. 6+ yrs	+1.55	0.27	0.12
3-6 yrs vs. 6+ yrs	+1.48	0.28	0.14
CD4 change			
CD4 stable vs. CD4 change	+0.59	0.43	0.56
HIV RNA viral load			
Viral load stable vs. viral load change	+1.29	0.23	0.20
Illicit substance toxicity (<i>n</i> = 86)			
Tox stable vs. Tox change	+0.67	0.26	0.50
sRCS scores			
Cognitively stable vs. improvement	-0.28	0.34	0.78
Cognitively stable vs. decline	-0.80	0.26	0.43
Cognitive decline vs. improvement	+0.31	0.37	0.76

Note. SEM = standard error of the difference between correlations; sRCS = standard regression-based change score; *N* = 88.

* two tailed *p* > .05.