

2019

A Comparative Study of the NAART and WRAT4 Word Reading Subtest to Estimate Reading Level

Elizabeth B. Campbell
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all



Part of the [Psychology Commons](#)

Repository Citation

Campbell, Elizabeth B., "A Comparative Study of the NAART and WRAT4 Word Reading Subtest to Estimate Reading Level" (2019). *Browse all Theses and Dissertations*. 2116.
https://corescholar.libraries.wright.edu/etd_all/2116

This Dissertation is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

**A COMPARATIVE STUDY OF THE NAART AND WRAT4 WORD READING
SUBTEST TO ESTIMATE READING LEVEL**

PROFESSIONAL DISSERTATION

SUBMITTED TO THE FACULTY

OF

**THE SCHOOL OF PROFESSIONAL PSYCHOLOGY
WRIGHT STATE UNIVERSITY**

BY

ELIZABETH CAMPBELL, PSY.M.

**IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PSYCHOLOGY**

Dayton, Ohio

July, 2020

COMMITTEE CHAIR: Gokce Ergun, Ph.D.

Committee Member: LaTrelle Jackson, Ph.D., CCFC, ABPP

Committee Member: Hyejin Kim, Ph.D.

Copyright by
Elizabeth Campbell
2019
All rights reserved

Abstract

Clients are often asked to fill out paperwork in medical settings, but varying reading levels can affect the ability to self-report. By screening for literacy level, clinicians can prevent potential patient confusion, frustration, and embarrassment. Clinicians can also avoid problems leading to misdiagnosis and providing materials that patients will not be able to understand or follow. The Wide Range Achievement Test – Fourth Edition (WRAT4) Word Reading subtest provides an estimation of literacy level as well as an estimation of premorbid ability. The North American Adult Reading Test (NAART) only provides an estimation of premorbid abilities, but it is quick to administer and available in the public domain. By correlating these word reading measures (NAART & WRAT4 Word Reading subtest), literacy level can be determined by using the NAART alone. This project was a systematic replication of a study conducted by Jones, Proto, Musso, Barker, and Gouvier (2012) using a demographically-different sample of participants. Results indicate that NAART error scores and WRAT4 Word Reading subtest raw scores were significantly correlated. Although the distributions of WRAT4 Word Reading subtest raw scores from the Jones et al. (2012) study and the current study significantly differed, there was not a statistically significant difference between the proportion of predicted reading levels below/at or above fourth or fifth grade and actual reading levels below/at or above fourth or fifth grade. This finding suggests that the original regression equation created by Jones et al. (2012) can be applied to the current sample to accurately predict reading classification.

Table of Contents

Abstract	iii
Table of Contents	iv
List of Tables	v
Acknowledgement	vi
Dedication	vii
Chapter 1. Statement of Problem	1
Chapter 2. Literature Review	4
Chapter 3. Method	25
Chapter 4. Results	30
Chapter 5. Discussion	35
References	44

List of Tables

Table 1. Demographic Statistics	26
Table 2. Descriptive Statistics	31
Table 3. Correlation Between NAART Error Scores and WRAT4 Word Reading Test Raw Scores	31
Table 4. Two-Sample Kolmogorov Smirnov Analysis Using WRAT4 Word Reading Test Raw Scores Between the Original and Current Study	32
Table 5. Jones et al. (2012) Regression Equation Predicting Number of People Above or Below Fourth Grade Reading Level	33
Table 6. Jones et al. (2012) Regression Equation Predicting Number of People Above or Below Fifth Grade Reading Level	33
Table 7. Significance of McNemar's Test for Predicting Reading Grade Levels Below/At or Above Fourth and Fifth Grade	34

Acknowledgement

I am very thankful to all of my family members and friends who have supported me during the last four years. I am grateful to Drs. Ergun, Allen, Jackson, Kim, Anspaugh, and Hurt for their input and encouragement. Finally, I am appreciative of all of the staff and faculty at Wright State School of Professional Psychology who helped me in this journey.

Dedication

I dedicate this dissertation to my father.

Chapter 1

Statement of the Problem

Clients are frequently asked to fill out paperwork in order to aid in medical treatment. Despite this request, many times there is no formal assessment given to determine the appropriateness of self-report measures based on required literacy level. By screening for literacy level before other assessments, clinicians can prevent potential patient confusion, frustration, and embarrassment. Presently, literacy measures are not administered to every client and there are only a few measures that provide this information. However, administering a quick measure of literacy would greatly benefit the interaction of health care professionals and clients.

This study is designed to benefit mental health professionals who wish to estimate reading literacy level with a reading test available in the public domain. It is hoped that the results of this study will aid professionals who plan on administering the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Hathaway, McKinley, & MMPI Restandardization Committee, 1989) or Personality Assessment Inventory (PAI; Morey, 1991) to quickly and cost effectively measure reading levels for clients in the older age range. Additional research is required to ensure that clients of any age can be assessed for reading level quickly and reliably in order to ensure comprehension of other assessments. The current study will also expand the field's knowledge regarding the use the North American Adult Reading Test (NAART) as a measure of reading level.

Aim and Purpose

The Wide Range Achievement Test – Fourth Edition (WRAT4) Word Reading subtest provides an estimation of literacy level as well as an estimation of premorbid intellectual ability. At this time, the NAART only provides an estimation of premorbid abilities. Because both tests produce an estimate of premorbid ability in the same format, they have been linked together in the literature. The NAART has become a popular assessment to give to clients due to its speed, ease of administration, and availability in the public domain. By correlating these two word reading measures (NAART & WRAT4 Word Reading subtest), literacy level can also be determined by using the NAART alone. One study has investigated this idea in the past, but more research is needed to justify using the NAART as a valid and reliable source of literacy level estimation.

This project is a systematic replication of a study conducted by Jones, Proto, Musso, Barker, and Gouvier (2012) using a demographically different sample of participants. Ages (primarily older), ethnicities (different proportion of African American and Caucasian), neurological statuses (healthy), education (greater range), and gender (primarily female) differ between the two samples.

The study will investigate the degree to which raw scores on the WRAT4 Word Reading subtest scores and NAART error scores vary together. The study will also implement the regression equation created in the original study to provide an estimated score on the WRAT Word Reading subtest based on NAART error scores alone. By verifying the accuracy of this estimation, mental health professionals will be able to administer the NAART as a stand-alone measure to obtain literacy level. Lastly, the results of the predictive equation will be compared to actual reading literacy level to

investigate the accuracy of the prediction (Jones et al., 2012). This comparison will aid researchers in understanding the generalizability of the predictive equation for different populations. The overall aim of this research is to provide data regarding the reliability of using the NAART to measure reading literacy level in older adults so that mental health providers can assess this variable quickly and easily to inform additional assessments.

Chapter 2

Literature Review

Clients are frequently asked to complete assessments, questionnaires, or general paperwork for medical and psychological appointments in order to inform assessment and treatment. These evaluations are sometimes lengthy and assume that each individual has a certain level of reading competence. Quick measures of word reading ability are often administered as part of a larger assessment, especially in neuropsychological and psychological assessment settings. Word reading tests are commonly used to predict premorbid abilities but can also be utilized to estimate current reading grade level.

Methods of Estimating Premorbid Ability

Clinicians completing assessments often require an estimate of a client's premorbid abilities, for the purpose of diagnostic clarity. Some disorders, such as major neurocognitive disorder, require a demonstrated decline in cognitive functioning in order to be diagnosed (Franzen, Burgess, & Smith-Seemiller, 1997). It is often the case that clinicians are not able to obtain historical information that would inform them about previous levels of functioning such as testing that took place prior to brain insult or injury. Therefore, measures aimed at predicting previous levels of functioning have been developed and modified. By obtaining this personal estimate, an individual's present performance can be compared to his or her previous abilities and more accurate diagnoses can be given (Lezak, Howieson, Loring, Hannay, & Fischer, 2004).

Demographics. The first assessment to estimate premorbid abilities based on demographic information alone was developed by Wilson et al. in 1978. This method was based on a regression equation and provided estimates for verbal intelligence quotient, performance intelligence quotient, and full-scale intelligence quotient. The authors found that education and race were the strongest predictors of intelligence quotient scores, but formulas were composed of weighted factors for age, sex, race, education, and occupation (Wilson et al., 1978).

More widely known than the formula created by Wilson et al. (1978) is the Barona formula that was developed later and in conjunction with the release of the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981). The Barona formula included additional demographic items to inform a more complex estimate. For example, the weighted factors of region of residence, urban or rural residence, and handedness were tested for significance in predictive abilities (Barona, Reynolds, & Chastain, 1984). The authors found education, race, and occupation to be the most powerful predictors of full-scale intelligence quotient in the newly developed equation. Several factors were not significantly predictive in relation to verbal, performance, or full-scale intelligence quotient. For example, handedness was not a significant predictive factor for any intelligence quotient score and was ultimately taken out of all equations.

Alternative equations have also been developed to adapt to other cultures such as United Kingdom (Crawford et al., 1989; Crawford & Allan, 1997). Equations that utilized age, education, and occupation were used to predict a significant amount of variance in intelligence quotient scores. Crawford and Allan (1997) found that occupation classification was the best predictor for verbal, performance, and full-scale intelligence

quotient. The authors also found that factoring in gender did not significantly add to predictive power. Similar to findings reported by Barona et al. (1984) utilizing an American sample of participants, equations for performance intelligence quotient were inferior to those predicting verbal or full-scale intelligence quotient in a British sample (Crawford & Allan, 1997).

Studies have shown mixed results with regard to the overall utility of the Barona formula and formulas that exclusively utilize demographic variables to predict premorbid intelligence. Eppinger, Craig, Adams, and Parsons (1987) found that the Barona formula overestimated intelligence quotient scores for healthy individuals when their intelligence quotient was less than 89. Additional research demonstrated that the Barona formula underestimates the intelligence quotient for those with a score above 110 (Ryan & Prifitera, 1990). Confirming the findings of the previous studies, several research studies have asserted that the Barona formula overestimated lower intelligence quotient scores and underestimated higher intelligence quotient scores on the WAIS (Goldstein, Gary, & Levin, 1986; Paolo & Ryan, 1992; Sweet, Moberg, & Tobian, 1990). Additionally, the categorization of occupation in the test has been criticized, as many careers do not clearly belong in a standard category (Eppinger et al., 1987), which can greatly impact a person's estimated intelligence. In a review of the limitations of the Barona formula, Hawkings (1995) asserted that the formula has a large standard error of estimation, is inaccurate when predicting scores in the upper and lower ranges and does not provide sufficient categorization of education or occupation.

Studies performed with brain-injured patients showed that predicted intelligence quotients scores using the Barona formula were significantly higher than actual

intelligence quotient scores (Eppinger et al., 1987). Neurologically healthy individuals did not display a significant difference in predicted versus actual scores. Seriousness of brain injury has been shown to create a larger discrepancy between predicted and actual intelligence quotient scores (Axelrod, Vanderploeg, & Rawlings, 1999). Severity of brain injury was determined by amount of time in coma; those with more time in coma had larger differences between predicted and actual intelligence quotient scores.

Researchers have been critical of tests of premorbid functioning that are based solely on demographic variables because they ignore other contributors to intelligence such as family support, economic status, and quality of education (Schinka & Vanderploeg, 2000; Strauss, Sherman, & Spreen, 2006). However, in a comparative study of methods of estimating premorbid abilities, Axelrod, Vanderploeg, and Schinka (1999) found that methods using only demographic data were equally as accurate as those that used other means of estimation such as performance methods. Contradictory findings highlight the discrepancies between studies in this area and indicate that additional research needs to be conducted.

Clinical judgment. Estimates of premorbid abilities made purely from clinical judgment are similarly rooted in demographically-based factors, but are evaluated subjectively (Kareken, Gur, & Saykin, 1995). Due to many limiting factors, judgments made purely from clinical evaluation are not common in neuropsychological practice. Pointing to the inaccuracies of human thought and how these can translate into imprecise premorbid predictions, Faust (1986) emphasized the limitations of human cognition and Kareken (1997) asserted that all clinicians carry judgment biases. Other criticisms of this subjective method have included clinician's tendency to overestimate cognitive decline in

the elderly, which may lead clinicians to overestimate premorbid abilities (Crawford, Millar, & Milne, 2001).

Upon reviewing available literature, Wedding and Faust (1989), as a deterrent to using mere clinical evaluation, reiterated the limitations of human judgment, but also suggested methods to improve decision making skills. Suggestions included clinical awareness that estimates are not perfectly accurate and necessary allowance for errors in judgment. The authors also recommended not overvaluing unique details of the case that may be easily remembered but are not necessarily predictive of other scores. Wedding and Faust (1989) also suggested that clinicians should avoid relying heavily on data presented at the beginning of an evaluation compared to data presented toward the end and attempt to actively seek evidence for other diagnoses. Overall, Wedding and Faust (1989) asserted that, whenever possible, avoiding decisions based purely on clinical judgment is best.

Several studies have suggested that statistical methods of estimating intelligence quotient scores are superior to clinical estimation (Dawes, Faust, & Meehl, 2002; Meehl, 1954; Wedding, 1983). In one study comparing clinical judgments to demographic formulas, Kareken and Williams (1994) found that although the two predictions of intelligence did not produce significantly different results, clinicians tend to have higher confidence in personal predictions. The researchers argue that limitations inherent in human reasoning and memory have contributed to inaccurate predictions of premorbid intelligence quotient and inflate levels of confidence in predictions (Kareken & Williams, 1994).

Best performance. Lezak, Howieson, Loring, Hannay, and Fischer (2004) first introduced the idea of the best performance method. In this estimation of premorbid functioning, intelligence quotient capacity is based on the highest level of cognitive performance in the client's lifetime. This method is a more comprehensive evaluation that considers test performance as well as personal history and observed behavior. The clinician should reflect on the client's current cognitive performance, observable behaviors, as well as premorbid achievements.

Similar to other methods of premorbid evaluation, the best performance method has been criticized. Mortensen, Gade, and Reinisch (1991) asserted that the best performance method is not indicative of general abilities. A client may be skilled in one area while lacking skill in another. Mortensen et al. (1991) also stated that the best performance method overestimates intelligence quotient scores for normally-functioning as well as neurologically-impaired individuals. The best performance method also fails to account for scatter within a client's performance across different domains (Matarazzo & Prifitera, 1989).

Hold – don't hold. This method of estimating premorbid intelligence looks at current performance on a task that is thought to be resistant to change due to insult or time. Hold - Don't Hold tests have typically been based on a subtest of a Wechsler intelligence test or a test of word reading ability (Franzen et al., 1997). Over time, researchers have found that word reading measures are more reliable in this method than WAIS subtests because they have been shown to be more resistant to brain insult (Franzen et al., 1997).

WAIS subtests. First introduced prior to many of the other methods for estimating intellectual abilities, the Hold – Don't Hold index was based on the idea that subtests such as Vocabulary from the WAIS are resistant to neurological impairments (McFie, 1975; Yates, 1956). Wechsler (1958) described a more complex calculation of intelligence where hold test scores were subtracted from other tests labeled as Don't Hold. Researchers have suggested using multiple tests or a single test in order to obtain estimates for this index. It has also been suggested that selection of WAIS subtests should be based on the type of insult so that hold tests are not based on skills that could be impaired (Franzen et al., 1997).

The Hold – Don't Hold index has faced many criticisms. Russell (1972) asserted that all WAIS subtests are impacted by neurological impairment, which would imply that no WAIS subtest is appropriate to use as a hold test. Similarly, when considering a specific disease process such as Alzheimer's, researchers have asserted that all WAIS subtests scores are impacted by this form of dementia (Larrabee, Lergen, & Levin, 1985). Others have suggested that the area of the brain impacted determines whether a WAIS subtest can be used as a hold test (Klesges, Wilkening, & Golden, 1981). In a study using a sample of people with traumatic brain injuries, Hoofien, Vakil, and Gilboa (2000) found that when comparing the hold – don't hold index to the best performance method, the best performance method displayed a more accurate prediction of premorbid functioning. Overall, researchers have shown that all WAIS subtests can be affected by brain insult.

Word reading. At the present time, word reading tests such as the NAART are commonly used to predict previous levels of intellectual functioning (Law & O'Carroll,

1998). First introduced by Nelson and McKenna (1975), word reading tests used to estimate a client's intelligence quotient have a strong basis in research, which established that clients with dementia maintained their word reading skill even in the presence of other forms of cognitive decline (Law & O'Carroll, 1998; Mortensen, Gade, & Reinisch, 1991; Nelson & Willison, 1982). Franzen et al. (1997) outlined the following four assumptions of word reading measures: intelligence is highly correlated with reading, reading ability is more preserved than performance ability on WAIS subtests, the ability to read irregular words has more predictive power than reading regular words, and word reading examines previous knowledge rather than current cognitive abilities.

Further analysis of the predictive power of word reading tests has been demonstrated in many studies. Generally, researchers have found a significant relationship between intelligence and reading ability in the neurologically-healthy population (Crawford et al., 1989). Additionally, reading tests have been shown to accurately predict intelligence quotient scores prior to brain injury (Crawford, Deary, Starr, and Whalley, 2001; Moss & Dowd, 1991). Lastly, the word reading performance of people afflicted with mild forms of dementia are mostly preserved (Crawford, Parker, Stewart, Besson, & Lacey, 1989; Fromm, Holland, Nebes, & Oakley, 1991).

Because word reading tests rely on crystallized intelligence as opposed to fluid intelligence, they are more resistant to effects from injury or aging. Crystallized intelligence consists of accumulated skills, experience, and knowledge (Horn & Cattell, 1967). This type of intelligence has been shown to decline only minimally in old age. Fluid intelligence, which consists of a person's ability to solve novel problems, declines dramatically with age (Schaie, 2005).

Berry et al. (1994) were the first to demonstrate that a reading test could correlate with intelligence testing performed in the past. The authors originally administered the WAIS-R and after three-and-a-half years they administered the North American Adult Reading Test – Revised (NAART; Blair & Spreen, 1989). A sample of neurologically-healthy older adults was used for the study and the researchers found similar predictive power to other studies where both the intelligence and reading tests are performed simultaneously. Another study performed by Crawford et al. (2001) analyzed the retrospective predictive power between performance on the National Adult Reading Test (NART; Nelson, 1982; Nelson & Willison, 1991) at age 77 and intelligence quotient scores at age 11. Researchers found that current NART scores significantly predicted intelligence quotient at age 11. This evidence further demonstrates that reading tests have the ability to predict past intelligence rather than simply reflecting current intellectual abilities.

However, overall level of intelligence is an additional factor to consider when predicting word reading test performance. Johnstone Callahan, Kapila, and Bouman (1996) examined the Wide Range Achievement Test-Revised (WRAT-R) Word Reading subtest and NAART performance across intelligence quotient levels. The authors found that both tests underestimated higher intelligence ranges and overestimated lower intelligence ranges. This finding suggests that both word reading tests are better able to accurately predict premorbid functioning for individuals of average intelligence (Johnstone et al., 1996).

Nelson and Willison (1982) developed the National Adult Reading Test (NART) to provide an accurate measure of premorbid functioning. The NART was created in

Britain and requires clients to read 50 phonetically irregular words aloud. Use of phonetically irregular words allows clinicians to assess for premorbid familiarity with the words rather than current ability to apply phonetic rules (Blair & Spreen, 1989). Premorbid familiarity with words is relatively preserved in people experiencing neurological disease. A test assessing for ability to apply phonetic rules would not be able to capture a client's skill level before the onset of degeneration because this skill would also be affected. The reading of short words is important to these predictions because the examinee is not required to analyze a complicated visual image, which could be compromised by brain injury. Because the words used in this test are also irregular, people will not be able to increase scores by phonological decoding or guessing. For these reasons, Nelson and O'Connell (1978) asserted that performance is not reliant on current abilities but is determined by previous knowledge.

NAART. Blair and Spreen (1989) later revised the NART for a North American population. The revised test (NAART) consists of 61 phonetically irregular words, which contain 35 words from the original version of the test. Subjects read each word aloud to and are graded on accuracy in pronunciation. Originally, this assessment of verbal intelligence was researched for people aged 20 to 70. The test also provides an estimate of premorbid intellectual functioning (Johnstone et al., 1996). When developing the NAART, Blair and Spreen (1989) used a sample of 66 participants from ages 18 to 49. The sample consisted of 64 Caucasian participants and the ethnicity of the remaining participants was not listed. Since its original development, the test has been researched in many studies for standardization purposes to include participants with a greater age range and ethnic diversity. For example, Uttl (2002) had a sample of 351 participants with an

age range of 18 to 91 who mostly identified as Caucasian but included 15 African American, 18 Asian, 6 Hispanic, and 8 participants with unknown ethnicities or ethnicities classified as other. The other category was not defined (Uttl, 2002).

Internal consistency reliability for the NAART was .935, while interscorer reliability was .99. Uttl (2002) found that interrater reliability was .93. Split-half reliability was reported as .93 and test-retest reliability was reported as .98. In an additional study, Blair and Spreen (1989) ruled out demographic variables including age, sex, education, occupation, and country of origin as factors accounting for a significant amount of additional variance in intelligence scores. The NAART score was the only significant predictor of an individual's intelligence quotient (Blair & Spreen, 1989).

In another study aimed at establishing reliability and validity of the NAART, Uttl (2002) had a sample of 351 healthy adults between the ages of 18 to 91. Validity between the NAART and WAIS-R Vocabulary was .75. Uttl (2002) also investigated reliability and validity when divided by age group and found that the NAART measures all age groups with comparable accuracy. Correlation with full-scale intelligence on the Wechsler Adult of Intelligence Scale- Revised was .75 (Blair & Spreen, 1989) and the researchers ruled out demographic variables including age, sex, education, occupation, and country of origin as factors accounting for a significant amount of additional variance in intelligence scores. Uttl (2002) found that the reliability of the NAART with WAIS-R Vocabulary scores was .93.

WRAT4 Word Reading subtest. The WRAT4 Word Reading subtest (Wide Range Achievement Test – Fourth Edition [WRAT4; Wilkinson & Robertson, 2006] Word Reading subtest) is a measure that serves a similar function as the NAART. The

WRAT4 test was originally developed to measure academic skill in the areas of reading, sentence comprehension, spelling, and math computation. There are two equivalent forms of the test available when re-administration is needed. The Word Reading subtest measures literacy level through letter and word decoding (Wilkinson & Robertson, 2006). Clients identify letters and read words aloud from a stimulus sheet containing 70 items. This measure can be used for people between the ages of 5 to 94. The Word Reading subtest is traditionally combined with the Sentence Comprehension subtest to create a reading composite score, which gives an estimate of overall reading achievement.

With a norming sample size of 3,007 individuals, the WRAT4 contains normative data for a total of 19 age groups. All age groups were matched by education level, race/ethnicity, gender, and geographic location. Four race/ethnicity categories were created for African American, Caucasian, Hispanic, and others. The other ethnicity category was not defined in the study. Educational attainment was also divided into four categories consisting of less than high school diploma, high school diploma, some college, and college graduate. Regions were divided into West, North Central, South, and Northeast. Researchers indicated that students with disabilities pertaining to education were included in the sample but represented at a lower percentage than their presence in the real population (Wilkinson & Robertson, 2006).

Overall reliability of the WRAT4 is excellent with the Word Reading subtest exhibiting split-half reliability of .98. Internal consistency reliability for the Word Reading subtest has a median of .92, but this measure varies by age group (Wilkinson & Robertson, 2006). Internal consistency within subtests was measured as ranging from .87 to .93. Alternate form retest reliability was measured as .90 for the Word Reading subtest

(Wilkinson & Robertson, 2006). When considering external validity, the Word Reading subtest shows a median correlation of .71 with similar word reading subtests (Wechsler Individual Achievement Test – Second Edition Word Reading [The Psychological Corporation, 2002], Wechsler Individual Achievement Test – Second Edition Decoding [The Psychological Corporation, 2002], Woodcock-Johnson III Basic Reading [Woodcock, McGrew, & Mather, 2001], and Kaufman Test of Educational Achievement Second Edition Comprehensive Letter/Word Recognition [Kaufman & Kaufman, 2004]).

Utility of Word Reading Tests

Both the NAART and WRAT4 Word Reading subtest are commonly used to predict premorbid intelligence, although the original purpose of the WRAT4 Word Reading subtest was to establish a client's literacy level (Blair & Spreen, 1989; Wilkinson & Robertson, 2006). Determining literacy level can aid in establishing the appropriateness of other assessments. It is important for clinicians to obtain an estimate of reading ability when administering a battery of tests, especially when self-report measures are given. For example, the Minnesota Multiphasic Personality Inventory (MMPI) requires a fifth-grade reading level (Hathaway, McKinley, & MMPI Restandardization Committee, 1989) and the Personality Assessment Inventory (PAI) requires a fourth-grade reading level (Morey, 1991) in order to be completed in a valid manner.

While some word reading tests are capable of providing measures of literacy and premorbid ability estimates (Wilkinson & Robertson, 2006), other popular word reading measures only offer estimates of premorbid ability (The North American Adult Reading Test [NAART; Blair & Spreen, 1989]). If these two measures have a strong relationship,

clinicians could use the NAART alone to find an individual's premorbid level of functioning and to determine the appropriateness of other assessments in a given battery based on literacy level. The NAART is available in the public domain and has similar reliability and validity to the WRAT4 Word Reading subtest, making it preferable due to cost considerations. By obtaining a reading level for each client using this free measure, overall test load would be reduced and problems such as test fatigue, frustration, and potential invalidity could be avoided. When clients are given tests or paperwork they cannot read, acquired data is not useful to the clinician. Additionally, clients could feel overwhelmed by the expectation to read all materials.

In a study done by Davis, Crouch, Wills, Miller, and Abdehou (1990), researchers found a gap of at least five grade levels between average patient reading ability and patient education materials. In a healthcare setting, these discrepancies can introduce dangerous consequences when patients are given resources pertaining to medications or recommended health behaviors. The authors of this study also asserted that standard consent forms used in health care settings often require reading ability at the college level. In a similar study of health information reading level, Adkins and Singh (2001) found materials given out in ten outpatient psychiatric clinics had reading levels ranging from grade level 11.1 to 15.5. The authors called for reform in how patient material are written given that many patients in the United States are illiterate.

In order to avoid these issues, clinicians can administer a brief test of reading literacy level prior to other assessments. Many factors may contribute to why a clinician chooses one measure over another. The NAART is available in the public domain, which makes it more appealing for clinicians when considering test administration cost. Given

that the NAART and WRAT4 Word Reading subtest are relatively equivalent in terms of producing stable and consistent results, they can be used interchangeably. Few studies have examined the relationship between the WRAT4 Word Reading subtest and the NAART.

Systematic Replication of Previous Research

Jones, Proto, Musso, Barker, and Gouvier (2012) began investigating the prediction of literacy level from NAART administration by correlating the two measures (NAART and WRAT4 Word Reading subtest) and creating a predictive reading grade level from NAART error scores alone. After establishing a strong correlation, the authors created a linear regression equation to predict WRAT4 Word Reading raw scores from NAART error scores.

Both the NAART and WRAT4 Word Reading subtests were administered to each participant as part of a larger clinical screening for intake assessments in the clinic. Participants in the Jones et al. (2012) study consisted of 173 patients at a public HIV clinic in the southern region of the United States. The authors suggested that patients at this clinic often lacked insurance or did not have sufficient insurance coverage. The sample was primarily African American (90.8%). Other ethnicities included Caucasian (8.1%) and Asian (0.6%). Male participants accounted for 55% of the sample and most participants were middle aged. Participants ranged in age from 18 to 67 with a mean of 37.9 years old. Forty-four percent of participants did not complete high school, 37% were high school graduates, and another 19% reported education past high school. All participants were HIV positive. Clinical screening assessments took approximately 45 minutes to complete and were administered by a master's-level clinician supervised by

licensed clinical psychologists. Administration and scoring followed standardized procedure.

Final analysis of the data included 167 participants, excluding six people who represented outliers. Most of the outliers did not pronounce any words correctly on the NAART, demonstrating a floor effect. No significant differences were found when comparing African American participants to non-African American participants on age, education, NAART error scores, or WRAT4 Word Reading subtest raw scores. The correlation between NAART error scores and WRAT4 Word Reading subtest raw scores was significant, $r=-.87$, $p<.001$. Authors created a linear regression equation in order to calculate the WRAT4 Word Reading subtest raw score from the NAART error score alone. The linear regression equation was significant. The authors created a table to display the number of errors made on the NAART and corresponding reading grade level. When employing these results, clinicians can be 95% confident that the calculated reading level is within one grade level of what the participant would score on the WRAT4 Word Reading subtest (Jones et al., 2012).

Jones et al. (2012) wanted to investigate the relationship between the NAART and WRAT4 Word Reading subtest in an African American population and among participants with lower levels of education. Both of these populations are generally underrepresented in the neuropsychological literature. The researchers also wished to create a way to estimate reading level from NAART error scores alone. Jones et al. recognized that no participant committed less than 17 errors on the NAART; therefore, interpretation of results for individuals who miss less than 17 should be made cautiously. However, 17 or fewer errors would be classified in the highest reading level attainable

(13+). The highest reading ability represents a literacy level at or above 13 years of education. Specific distinctions cannot be made between participants who make 17 errors or less; however, reading level is assumed to be sufficient for most assessments because it is greater than 13 years of education. This study did not set strict standards on acceptable medical conditions for participants. The authors proposed that underlying neurological disorders could have impacted results. They also cautioned against generalizing this information to groups of people who are not African American and who have obtained 12 years of education or more.

Overall, the Jones et al. (2012) study set out to obtain an accurate reading level from the administration of the NAART alone. By correlating error scores from the NAART and raw scores from the WRAT4 Word Reading subtest, the authors were able to create a regression equation. Clinicians can be 95% confident that the calculated reading level based on the regression equation is within one grade level of what the participant would score on the WRAT4 Word Reading subtest. The sample in this study consisted of mostly African American males who were HIV positive impacting the generalizability of the results.

The Current Study

Correlating the NAART and WRAT4 Word Reading subtest is important for obtaining easily accessible literacy scores for patient populations. Further, research is needed to support accurate grade level estimations using NAART error scores for the general population. The current study will investigate the relationship between these two word reading tests and provide additional information on how these tests are related. This research will attempt to systematically replicate the Jones et al. (2012) study while using

participants with differing demographics such as ages (primarily older), ethnicities (different proportion of African American and Caucasian), neurological statuses (healthy), education (greater range), and gender (primarily female). The regression equation created in the original study will be used to estimate reading level of the current study's sample and compared to actual reading level. Analysis will be performed to measure the level of agreement between estimated and actual reading level using reading level requirements from two popular personality inventories.

Age. When considering cognitive changes across the lifespan, many people associate old age with sharp decline in all skills. However, some processes remain relatively stable such as autobiographical memory, long-term memory, semantic knowledge, and emotional processing. Crystallized intelligence including semantic memory and knowledge remain stable until late in life (Hedden & Gabrieli, 2004). Steady decline in ability has been demonstrated in the following processes across the lifespan: the formation of new memories, working memory, and executive functioning skills such as processing speed. However, disease often accompanies aging and some decline in proficiency cannot be blamed on aging as a factor independent of other influences such as dementia or hypertension.

Several studies have focused on variability of word reading test performance when considering the demographic variable of age. For example, Uttl (2002) asserted that NAART performance improves across the adult life span. Two studies found that older groups of participants ranging from age 60 to 79 scored 0.5 to 0.7 standard deviations higher than younger groups ranging from age 20 to 39 on the NAART (Graf & Uttl,

1995; Uttl & Graf, 1997). Further analysis across age groups found comparable reliability and validity as well as lack of bias for any age group.

Ethnicity. Researchers have linked ethnicity and education when speaking about word reading performance. Manly, Jacobs, Touradgji, Small, & Stern (2002) posited that educational experiences were generally dissimilar between majority and minority populations. Thus, years of education is not generally indicative of quality of education or word reading level. Evidence for this idea was found by Manly et al. (1998), which found that African Americans are more likely than White counterparts to be misdiagnosed as cognitively impaired when measures are used that take years of education into account.

Neurological status. With regard to WRAT4 Word Reading subtest performance across groups, Casaletto et al. (2014) found no difference in performance between HIV seropositive (HIV+) individuals and HIV seronegative (HIV-) groups. Although HIV+ individuals exhibited lower overall neuropsychological test performances on measures of processing speed, memory, and executive functioning, word reading performance was unaffected. Because the Jones et al. (2012) study was comprised of HIV+ individuals, Casaletto et al.'s findings suggest that further studies would find no difference in word reading performance based on the presence of HIV.

Education. With regard to education, Uttl (2002) found that NAART scores increased with higher education levels. This study established that scores on the NAART improve approximately 1.5 points per year of additional education. Many studies have focused on the difference between years of education and quality of education impacting neuropsychological assessment results. Johnstone et al. (1997) asserted that years of

education is not an adequate indicator of premorbid abilities because quality of education variables differ greatly between individuals. These authors proposed using word reading tests instead of obtaining a self-report of education level for a more accurate estimation of premorbid ability.

Gender. Much research has suggested that gender does not impact word reading scores. This has been shown in studies utilizing the WRAT4 Word Reading subtest (Berg, Durant, Banks, & Miller, 2016) as well as the NAART (Uttl, 2002). This research analyzed data for variance in scores between demographically divergent groups such as gender and did not find evidence to support significant differences.

Overall, studies have been mixed with regard to whether variables such as age, ethnicity, neurological status, education, and gender can impact word reading test performance. However, most suggest improvement in test scores based on the demographic differences in the current sample when compared to the Jones et al. (2012) sample. Scores are likely to vary based on age, level of education, or quality of education. As people age, crystallized intelligence scores, such as vocabulary knowledge, naturally increase (Horn & Catell, 1967). Ricketts, Nation, & Bishop (2007) found that as vocabulary scores increase, ability to read phonetically irregular words also increases. Therefore, increases in age potentially relate to an increase in crystallized intelligence scores as well as word reading ability for phonetically irregular words. Level of education or quality of education also increases intelligence scores (Balke-Aurell, 1982). Therefore, the current study is predicted to produce results that vary from the findings in Jones et al. (2012) due to demographic differences in the participant sample.

This project is a systematic replication of a study conducted by Jones et al. (2012) using a demographically different sample of participants. The study will investigate the degree to which raw scores on the WRAT4 Word Reading subtest scores and NAART error scores vary together. Lastly, the predictive equation given in previous research will be applied to the current study's sample to test accuracy of prediction using reading grade level cut-offs required for two well-known personality measures (Jones et al., 2012). Reading grade level cut-offs were determined by reading level requirements for two popular personality inventories. This will aid researchers in understanding the generalizability of the predictive equation in mental health settings for differing populations. The overall aim of this research is to provide data regarding the reliability of using the NAART to measure reading literacy level in older adults so that mental health providers can assess this variable quickly and easily to inform additional assessments.

First research hypothesis. Raw scores on the WRAT4 Word Reading subtest will correlate significantly with error scores on the NAART.

Second research hypothesis. The distribution of WRAT4 Word Reading scores will significantly differ between the samples from the original study and the current study.

Third research hypothesis. The regression equation created by Jones et al. (2012) will accurately classify if reading level falls at/above or below the fourth or fifth grade for the current study's sample.

Chapter 3

Method

Participants

This study used archival data from the former Community Memory Clinic (CMC) at the Ellis Human Development Institute in Dayton, Ohio. The director of the CMC provided authorization to review and analyze the data. The sample was recruited from community mental health centers in Dayton, Ohio and the surrounding area. Approval to conduct the study was obtained from a university-based panel that assesses the appropriateness and ethics of the study for participation by human subjects. Participating facilities included Victor J. Cassano Community Health Center, Charles R. Drew Health Center, St. Leonard's Hospital, and Robert A. Vogel Health Center. Participants had no prior diagnoses of psychiatric or neurological disorders. A total of 85 participants met inclusion criteria for this study. Participants were included in the study if they were neurologically healthy and age 55 or above. Data regarding history of neurological problems was gathered through a demographic questionnaire that was completed by each participant. Exclusionary criteria included previous self-reported diagnoses of dementia, seizure disorder, head trauma, attention deficit disorder, or stroke.

Participants ranged in age from 55 to 92. Most participants fell between the ages of 60 and 79 (54.1%). About half (55.29%) identified as Black/African American and 37.65% identified as White/non-Hispanic. Education level ranged from 7 to 20 years, with 12 years of education representing the mode (49.41%). Other participants had 16

years (20%) and 14 years (7.06%) of education. The participants consisted mostly of females (71.76%). See Table 1 for a list of all demographic statistics.

Table 1

Demographic Statistics

		Frequency	Percentage
Age	55-59	17	20.00
	60-69	23	27.05
	70-79	23	27.05
	80-89	21	24.71
	90 +	1	1.17
Race/Ethnicity	American Indian/Alaskan Native	1	1.18
	Asian	2	2.35
	Native Hawaiian/Pacific Islander	0	0.00
	Black/African American	47	55.29
	White	32	37.65
	Multiple Ethnicities	3	3.53
Type of Degree	None	3	3.53
	High School Diploma	42	49.41
	GED	3	3.53
	Associates Degree	6	7.06
	Bachelors Degree	17	20.00
	Masters Degree	12	14.12
	Professional Degree	0	0.00
	Doctoral Degree	1	1.18
	Unreported	1	1.18
Gender	Male	24	28.24
	Female	61	71.76

Note. $N = 85$

Materials

Reading literacy level was measured by the Wide Range Achievement Test-Fourth Edition (WRAT4; Wilkinson & Robertson, 2006). The North American Adult

Reading Test (NAART; Blair & Spreen, 1989) was also administered to all participants. The NAART is typically used to predict premorbid cognitive abilities. Reliability and validity information for each assessment can be found in the literature review.

Wide Range Achievement Test- Fourth edition (WRAT4). The WRAT4 is an achievement measure that provides estimates of word reading, spelling, sentence comprehension, and math computation. Wilkinson (2006) standardized the subtests on a sample of 3000 people who were 5-94 years old. The word reading subtest measures reading ability through letter and word identification, and it is the only subtest from the WRAT4 used in this study. Raw score is calculated by adding the number of correctly pronounced words out of a total of 55 words. Raw scores are typically converted into standard scores according to age.

North American Adult Reading Test (NAART). Blair and Spreen (1989) developed the NAART as a measure of premorbid ability based on reading fluency. It presents phonetically irregular words and participant's reading ability relies on premorbid familiarity with the words rather than current ability to apply phonetic rules. Total number of errors in pronunciation are added and used to calculate a predicted full-scale intelligence quotient.

Procedure

The current research utilized data from the Allen (2009) study, which was entered into a data file. The original research study was approved by the Wright State University Institutional Review Board, and participants underwent testing for cognitive, adaptive, and word reading ability to assess the impact of education on estimates of premorbid intelligence. Administered measures were completed over two sessions. Demographic

information was collected in interviews with participants and caregivers. Work, education, and medical history were collected through a demographic questionnaire.

The current study was a systematic replication study of research conducted by Jones et al. (2012) with a demographically different subject population. It analyzed the WRAT4 Word Reading subtest raw score and NAART error score by entering data into SPSS from the Allen (2009) study. A new file was created in SPSS. NAART error scores and WRAT4 Word Reading subtest raw scores were tested for significant correlation. Then, the two samples of WRAT4 Word Reading scores (original and current study) were compared using a Two-Sample Kolmogorov-Smirnoff test to assess for significant differences between the distributions. Lastly, McNemar's test was run using the Jones et al. (2012) regression equation to predict reading level of the sample from this study. P-values were analyzed to determine whether predicted reading level correctly classified whether reading level fell above or below fourth or fifth grade reading level.

Design

Statistical analysis included a Pearson correlation analysis of NAART error scores and WRAT4 Word Reading raw scores to examine the relationship between WRAT4 Word Reading subtest raw scores and NAART error scores. Assumptions for this type of analysis include that data is paired, randomly sampled, and quantitative, scatterplot of the data conforms to a linear pattern, and any outliers are removed (Triola, 2014).

To examine the second hypothesis, a two-sample Kolmogorov Smirnoff test was run to determine whether the distributions of WRAT4 Word Reading scores from the original and current study significantly differ. Assumptions for the second analysis

include that observations in the two samples are randomly selected and independent of one another with at least an ordinal scale of measurement (Sheskin, 2011).

The final hypothesis, which compared the estimated reading level using the regression equation from Jones et al. (2012) and actual reading level, were tested using McNemar's test. Each participant in the sample was classified as having an estimated reading level above or below the fourth and fifth grade. This classification was compared to the participant's actual reading level to test for significant difference. Assumptions for a McNemar's test include that the data came from a random sample of a population, each observation is independent of others, scores are a dichotomous and categorical measure involving two mutually exclusive categories, and the sample was not extremely small in size (Sheskin, 2011). Lastly, a Cohen's kappa analysis was run to analyze if the prediction in reading level provided by the Jones et al. (2012) regression equation added to prediction above and beyond levels expected by chance. Assumptions for a Cohen's kappa test include that the data is measured categorically, the same two raters are used for each observation, and the data is rated independently (Fleiss, Levin, & Palik, 2003).

Chapter 4

Results

The current study focused on applying a previously-generated regression equation to estimate reading level in a sample that is demographically different than the original sample. The results presented below are based on the three research questions of the study.

First Research Question

A Pearson correlation was run in order to determine the relationship between NAART error scores and WRAT4 Word Reading Test raw scores. Assumptions for this type of analysis include that data is paired, randomly sampled, and quantitative, scatterplot of the data conforms to a linear pattern, and any outliers are removed (Triola, 2014). The sample of data from the current study met these requirements. The means and standard deviations of the NAART error scores ($M = 30.47$; $SD = 13.93$) and the WRAT4 Word Reading Test raw scores ($M = 57.48$; $SD = 8.84$) are shown in Table 2. The corresponding ranges are also presented in Table 2. The results from Table 3 indicated that NAART error scores and WRAT4 Word Reading Test raw scores are significantly correlated. In the current study's sample, NAART error scores were significantly negatively related to WRAT4 Word Reading Test raw scores ($r = -.91$).

Table 2

Descriptive Statistics

	Mean	Standard Deviation	Minimum	Maximum
NAART error scores	30.47	13.93	5.00	58.00
WRAT4 Word Reading Test raw scores	57.48	8.84	32.00	70.00

Note. $N = 85$

Table 3

Correlation Between NAART Error Scores and WRAT4 Word Reading Test Raw Scores

	WRAT4 Word Reading Test Raw Scores
NAART error scores	-.91*

Note. $N = 85$

* $p < .001$

Second Research Question

A Two-Sample Kolmogorov-Smirnov Test was run in order to test whether the distribution of WRAT4 Word Reading Test raw scores in the original Jones et al. (2012) study significantly differed from the WRAT4 Word Reading Test raw scores in the current study. Table 4 shows that the distributions for the two samples significantly differ ($Z = 3.804, p < .001$). Assumptions were met for this analysis and included that observations in the two samples are randomly selected and independent of one another with at least an ordinal scale of measurement (Sheskin, 2011).

Table 4

Two-Sample Kolmogorov Smirnov Analysis Using WRAT4 Word Reading Test Raw Scores Between the Original and Current Study

Most Extreme Difference (Absolute)	0.51
Kolmogorov-Smirnov Z	3.80
Asymp. Sig. (2-tailed)	<0.001

Note. $N = 85$

Third Research Question

McNemar’s Test was run to analyze the proportion of participants that were correctly or incorrectly classified as above or below the fourth- and fifth-grade reading level when the regression equation from the Jones et al. (2012) study was applied to the current study’s data set. This analysis accounted for the predicted reading level generated from the original regression equation and actual reading level as measured in the current study. Table 5 displays the number of people in the study correctly ($n = 83$) and incorrectly classified as above or below fourth-grade reading level ($n = 2$). Table 6 displays the number of people in the study correctly ($n = 83$) and incorrectly classified ($n = 2$) as above or below fifth-grade reading level. Assumptions for a McNemar’s test were met for this sample and include that the data came from a random sample of a population, each observation is independent of others, scores are a dichotomous and categorical measure involving two mutually exclusive categories, and the sample was not extremely small in size (Sheskin, 2011).

Table 5

Jones et al. (2012) Regression Equation Predicting Number of People Above or Below Fourth-Grade Reading Level

	Predicted Below Fourth	Predicted Fourth or Above
Actual Below Fourth	0	2
Actual Fourth or Above	0	83

Note. N = 85

Table 6

Jones et al. (2012) Regression Equation Predicting Number of People Above or Below Fifth-Grade Reading Level

	Predicted Below Fifth	Predicted Fifth or Above
Actual Below Fifth	4	1
Actual Fifth or Above	1	79

Note. N = 85

McNemar's test determined that there was not a statistically significant difference in the proportion of predicted reading levels below/at or above fourth grade and the actual reading levels below/at or above fourth grade ($\chi^2 = .50$). Similarly, the McNemar's test determined that there was not a statistically significant difference in the proportion of predicted reading levels below/at or above fifth grade and the actual reading levels below/at or above fifth grade ($\chi^2 = .99$). These results are presented in Table 7. Raw level of agreement using McNemar's test when examining fourth-grade reading level is 97.6%.

However, the calculated level of agreement by chance is also 97.6%. Therefore, the

applied regression equation does not add to predicted reading level above chance levels. Consequently, Cohen's Kappa for this analysis was .00. Raw level of agreement using McNemar's test when examining fifth-grade reading level is 97.6%. Calculated level of agreement by chance for fifth-grade reading level is 88.93%. Cohen's Kappa for this analysis was .78, representing a moderate level of agreement between the prediction of reading level by the Jones et al. (2012) regression equation and the actual reading level in the current study's sample while accounting for agreement by chance. Assumptions were also met for carrying out a Cohen's kappa, which include that the data is measured categorically, the same two raters are used for each observation, and the data is rated independently (Fleiss, Levin, & Palik, 2003).

Table 7

Significance of McNemar's Test for Predicting Reading Grade Levels Below/At or Above Fourth and Fifth Grade

	Below/At or Above Fourth Grade	Below/At or Above Fifth Grade
Significance (2-tailed)	.50	.99
Cohen's Kappa	.00	.78

Note. $N = 85$

Chapter 5

Discussion

The principal goals of this research study were to expand upon current research concerning the relationship between the WRAT4 Word Reading subtest and the NAART and to examine the generalizability of a previously-created regression equation linking the two tests in a demographically-different sample of participants. The first research question examined the correlation between the raw scores on the WRAT4 Word Reading subtest and the error scores on the NAART. Similar to other studies that have examined this relationship, the current study found that the WRAT4 Word Reading subtest and error scores on the NAART are significantly correlated. For example, Johnstone et al. (1996) found a strong relationship between the assessments ($r = .87$) when comparing WRAT4 Word Reading subtest scores and Verbal intelligence quotient estimates that were found utilizing NAART error scores. Additionally, Jones et al. (2012) found a similar association ($r = -.87, p < .001$) when examining raw scores on the WRAT4 Word Reading subtest and NAART error scores.

Supporting evidence that these two assessments should be strongly correlated has been found in validation studies as well. Blair and Spreen (1989) found that validity between the NAART and WAIS-R Vocabulary was .75. Correlation with full-scale intelligence on the Wechsler Adult of Intelligence Scale- Revised was also .75 (Blair & Spreen, 1989). When considering external validity, the WRAT4 Word Reading subtest shows a median correlation of .71 with similar word reading subtests (Wechsler

Individual Achievement Test – Second Edition Word Reading [The Psychological Corporation, 2002], Wechsler Individual Achievement Test – Second Edition Decoding [The Psychological Corporation, 2002], Woodcock-Johnson III Basic Reading [Woodcock, McGrew, & Mather, 2001], and Kaufman Test of Educational Achievement Second Edition Comprehensive Letter/Word Recognition [Kaufman & Kaufman, 2004]). The current study examined a sample of participants that differed from the aforementioned studies in that the sample was neurologically healthy, primarily elderly, and more highly educated. Results showed that the two tests are highly correlated in the current sample ($r = -.91$, $p < .001$) as well. Therefore, it appears as though these two assessments examine a similar construct and the current study has expanded upon past research linking the two tests with a demographically-different population.

The second research question examined the distribution of scores between the Jones et al. (2012) study and the current study. A Two-Sample Kolmogorov-Smirnov Test was carried out and found that the distribution of WRAT4 Word Reading subtest raw scores significantly differ between the original and current study. This finding affirms the conclusion that it would be difficult to generalize findings from the Jones et al. (2012) sample to the sample from the current study because the distributions of WRAT4 Word Reading raw scores differ.

Prior research offers insight into why the two distributions were significantly different. Participants in the current study differed in the following demographic variables from participants in the Jones et al. (2012) study: age (primarily older), education (greater range), and ethnicity (equal proportion of African American and Caucasian). Uttl (2002) asserted that NAART performance improves across the adult life

span. Two additional studies found that older groups of participants ranging from age 60 to 79 scored 0.5 to 0.7 standard deviations higher than younger groups ranging from age 20 to 39 on the NAART (Graf & Uttl, 1995; Uttl & Graf, 1997).

With regard to education, Uttl (2002) found that NAART scores increased with higher education levels. This study established that scores on the NAART improve approximately 1.5 points per year of additional education. However, many studies have focused on the difference between years of education and quality of education impacting neuropsychological assessment results. When considering ethnicity, Manly et al. (2002) posited that educational experiences are generally dissimilar between majority and minority populations. Dotson, Kitner-Triolo, Evans, and Zonderman (2008) found that reading ability better predicts cognitive functioning than years of education for low and higher socioeconomic status African Americans and low socioeconomic status Caucasians. Although socioeconomic status was not part of the data collected in the current study, this finding links ethnicity and level of socioeconomic status to quality of education. It provides evidence that racial minorities and people from lower socioeconomic status backgrounds are disadvantaged with regard to quality of education. Taken together, the research suggests that samples which differ with regard to these three variables are likely to differ.

Because of the powerful impact of these variables on scores for word reading tests, it follows that it may be difficult for one regression equation to accurately predict reading level for the general population without accounting for these demographic differences in some way. However, using a statistical test that classifies reading level as above or below a certain grade level would be statistically easier to predict than obtaining

an individual reading level estimate with accuracy. Practically, word reading tests in mental health settings will most likely be used to inform the appropriateness of other measures such as personality assessments. Therefore, it is worthwhile to investigate the predictive ability of the Jones et al. (2012) regression equation to classify reading level above or below a certain grade level. This information would give the clinician all the information necessary to deem additional assessments as appropriate or inappropriate.

The last research question utilizes the regression equation created by Jones et al. (2012) to predict reading levels above or below fourth- and fifth-grade cut offs in the sample from the current study. This analysis was carried out to analyze the generalizability of the original regression equation for a different sample. Correct classification of reading levels at fourth and fifth grade were investigated due to the required reading levels of two well-known personality measures. The PAI (Morey, 1991) requires a fourth-grade reading level to complete while the MMPI-2 requires a fifth-grade reading level (Hathaway, McKinley, & MMPI Restandardization Committee, 1989). McNemar's test determined that there was not a statistically significant difference in the proportion of predicted reading levels below/at or above fourth or fifth grade and the actual reading levels below/at or above fourth or fifth grade. Although the distributions of WRAT4 Word Reading test raw scores differ significantly between the two samples as related to the second research question, a categorical test was able to correctly classify when reading levels would fall above or below a certain grade. Categorical data is easier to predict with accuracy than continuous data, resulting in similar predicted and actual reading level scores. This finding suggests that the original regression equation created by Jones et al. (2012) can be applied to the current sample to accurately predict if reading

grade level is above or below the fourth or fifth grade. Therefore, a new regression equation does not have to be created in order to predict whether or not an individual would be able to comprehend two personality measures used as guidelines for this study.

In analyzing this result, it is crucial to also consider the level of agreement expected by chance and whether the raw level of agreement seen in McNemar's test improves upon chance prediction. With regard to fourth grade reading level, the regression equation created by Jones et al. (2012) does not add any additional predictive value beyond chance levels. When considering fifth-grade reading level, the regression equation did add some value beyond agreement by chance. The difference in added predictive value between grade levels can likely be attributed to the fact that a large proportion of the sample was predicted and able to read at or above a fourth-grade level. In fact, the only agreement between predicted and actual fourth-grade reading level was seen in this scenario, meaning that agreement by chance was very high. When analyzing the McNemar's test for fifth grade, there are some cases where predicted and actual reading level are below the fifth grade, making agreement by chance lower. Because there is room for improvement in the prediction, McNemar's test was able to provide a better level of prediction than agreement by chance.

Limitations

Very few participants had reading levels below fourth or fifth grade in the current sample. Only two participants had reading levels below the fourth grade and only five participants had reading levels below the fifth grade. More than 36 million adults in the United States (approximately 11%) cannot read above a third-grade level (ProLiteracy, 2006). When considering basic literacy skills in the state of Ohio, a study conducted by

the United States Department of Education estimated that 9% of the population lacked basic literacy skills (National Center for Education Statistics, 2003). Therefore, results from the sample used in this study were not generally representative of the United States or the state of Ohio, as it would be expected that a larger portion of the sample would score lower on word reading measures.

Another limitation to consider involves selection bias. The Allen (2009) study did not keep a record of study drop-out rates or potential participants who refused to participate. Therefore, it is possible that potential participants with lower levels of reading ability self-selected out of the study because they were made aware that cognitive abilities would be tested. Additionally, results from the study are limited in that they not generalizable to the larger population and only apply to the sample tested. Furthermore, the current study is imperfect because the sample size was smaller than the sample size utilized in the Jones et al. (2012) study. Matching sample sizes would have made for an ideal comparison. Lastly, neither the NAART or WRAT4 Word Reading subtest should be used in isolation to make a diagnosis. The present study suggests only that these assessments are used to estimate reading level in mental health settings where reading level is a crucial consideration in patient interaction.

Future Directions and Implications

Future research could benefit from replicating the format of the current study with a sample that has more variability in reading levels for participants. It would also be useful to obtain a sample of participants with greater variance when considering level of education and socioeconomic status. This would allow researchers to test the Jones et al. (2012) regression equation in a sample that is less homogeneous. Results from a study of

this kind that support significant prediction using the Jones et al. (2012) regression equation could provide evidence for a more general use for the equation in mental health settings. Additionally, a sample size that is comparable to the sample size used to inform the original equation would be ideal for the best results.

Research that investigates fast, easy, and free estimation of adequate reading level for environments that value quick assessment is vital to providing the best patient care. This study was tailored to assess for accurate prediction of whether a patient would be able to comprehend personality measures in mental health settings. A study conducted by Adkins and Singh (2001) examined the required reading level of materials provided in ten outpatient psychiatric clinics. The researchers concluded that resources were written at reading levels ranging from 11.1 to 15.5, rendering them incomprehensible to a large proportion of their patients. Without adequate understanding of the content of provided materials, there could be many negative outcomes such as misdiagnosis. Additionally, psychologists in this situation may provide incomprehensible materials, write inappropriate recommendations, or provide therapy in an unsuitable modality simply because a patient is unable to understand the words being utilized. A mental health professional could also misinterpret lack of compliance with homework as uncooperative when a patient is simply unable to do the work because of an inability to read provided content.

Beyond exclusively therapeutic advantages, a free assessment of reading level could serve as a screener to inform additional administered assessments in a clinical setting. This type of measurement could assist clinicians in choosing the type of assessments given that would be most efficacious for patient diagnosis. Psychologists are

in constant communication with their patients through assessments, interventions, and forms that are all based on an assumption that the patient is able to comprehend the material.

In addition to assisting psychologists, this kind of free reading literacy measure has the potential to benefit multiple disciplines and settings in health care. Professionals in the fields of social work and medicine could benefit immensely from reading level estimations to inform patient care. For example, Davis et al. (1990) found that patient reading ability and reading level of patient materials differed at least five grade levels in five health clinics. In healthcare settings, these gaps can impact a patient's ability to follow through with appropriate health-related directives and have potentially serious consequences. These researchers also found that examined consent forms were written at a college reading level, implying that they are beyond the understanding of most of the patients that signed them. A screener for reading level could greatly benefit the treatment of those with lower levels of education or those who learned English as a second language.

Considering settings outside of health care, this kind of screener could also benefit community organizations, such as spiritual settings, where information is constantly exchanged between organization leaders and participants. For example, spiritual literature or verbally-delivered spiritual messages and the comprehension of these forms of communication is dependent upon level of understanding of the group receiving the message. By asking group members to participate in reading level screeners, leaders of organizations could communicate their concern for the

comprehension of the entire group as well as ensuring comprehension and effective communication.

The potential benefits from a quick, easy, and free screener to assess reading level are vast, but more research is needed to make this a viable option in health settings. A more solid foundation of evidence is required to ensure that patient reading level misclassifications are avoided and the best patient care is provided.

References

- Adkins, A. D., & Singh, N. N. (2001). Reading level and readability of patient education materials in mental health. *Journal of Child and Family Studies, 10*(1), 1-8. doi:10.1023/A:1016617115729
- Allen, J. B. (2009). [Impact of quality of education on cognitive status of African-American older individuals]. Unpublished raw data.
- Axelrod, B. N., Vanderploeg, R. D., & Rawlings, D. B. (1999). WAIS-R prediction equations inpatients with traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 21*, 17-26.
- Axelrod, B. N., Vanderploeg, R. D., & Schinka, J. A. (1999). Comparing methods for estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology, 14*(4), 341-346.
- Barona, A., Reynold, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology, 52*, 885-887.
- Berg, J., Durant, J., Banks, S. J., & Miller, J. B. (2016). Estimates of premorbid ability in a neurodegenerative disease clinic population: Comparing the Test of Premorbid Functioning and the Wide Range Achievement Test, 4th Edition. *The Clinical Neuropsychologist, 30*(4), 547-557. doi:10.1080/13854046.2016.1186224

- Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the national adult reading test. *The Clinical Neuropsychologist*, 3(2), 129-136.
doi:10.1080/13854048908403285
- Balke-Aurell, G. (1982). *Changes in ability as related to educational and occupational experience*. Goteborg, Sweden: Acta Universitatis Gothoburgensis.
- Berry, D. R., Carpenter, G. S., Campbell, D. A., Schmitt, F. A., Helton, K., & Lipke-Molby, T. (1994). The New Adult Reading Test-Revised: Accuracy in estimating WAIS-R IQ scores obtained 3.5 years earlier from normal older persons. *Archives of Clinical Neuropsychology*, 9(3), 239-250. doi:10.1016/0887-6177(94)90029-9
- Casaleto, K. B., Cattie, J., Franklin, D. R., Moore, D. J., Woods, S. P., Grant, I., & Heaton, R. K. (2014). The Wide Range Achievement Test–4 reading subtest 'holds' in HIV-infected individuals. *Journal of Clinical and Experimental Neuropsychology*, 36(9), 992-1001. doi:10.1080/13803395.2014.960370
- Crawford, J. R., & Allan, K. M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, 11(2), 192-197. doi:10.1080/13854049708407050
- Crawford, J. R., Deary, I. J., Starr, J., & Whalley, L. J. (2001). The NART as an index of prior intellectual functioning: A retrospective validity study covering a 66-year interval. *Psychological Medicine*, 31(3), 451-458.
doi:10.1017/S0033291701003634
- Crawford, J. R., Millar, J., & Milne, A. B. (2001). Estimating premorbid IQ from demographic variables: A comparison of a regression equation vs. clinical

judgment. *British Journal of Clinical Psychology*, 40(1), 97-105.

doi:10.1348/014466501163517

Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A., & Lacey, G. (1989).

Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology*, 28(3), 267-273.

doi:10.1111/j.2044-8260.1989.tb01376.x

Crawford, J. R., Stewart, L. E., Cochrane, R. H. B., Foulds, J. A., Besson, J. A. O., &

Parker, D. M. (1989). Estimating premorbid IQ from demographic variables: Regression equations derived from a UK sample. *British Journal of Clinical Psychology*, 28(3), 275-278.

Davis, T. C., Crouch, M. A., Wills, G., Miller, S., & Abdehou, D. M. (1990). The gap between patient reading comprehension and the readability of patient education materials. *The Journal of Family Practice*, 31(5), 533-538.

Dawes, R. M., Faust, D., & Meehl, P. E. (2002). Clinical versus actuarial judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 716-729). Cambridge, United Kingdom: Cambridge University Press.

Dell, C. A., Harrold, B., & Dell, T. (2008). Test Review: Wide Range Achievement Test Fourth Edition. *Rehabilitation Counseling Bulletin*, 52(1), 57-60.

Dotson, V. M., Kitner-Triolo, M., Evans, M. K., & Zonderman, A. B. (2008). Literacy based normative data for low socioeconomic status African Americans. *The Clinical Neuropsychologist*, 22(6), 989-1017. <https://doi.org.ezproxy.libraries.wright.edu/10.1080/13854040701679017>

- Eppinger, M. G., Craig, P. L., Adams, R. L., & Parsons, O. A. (1987). The WAIS-R index for estimating premorbid intelligence: Cross-validation and clinical utility. *Journal of Consulting and Clinical Psychology, 55*, 86-90.
- Faust, D. (1986). Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice, 17*(5), 420-430.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. Hoboken, NJ: John Wiley & Sons, Inc.
- Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology, 12*, 711-738.
- Fromm, D., Holland, A. L., Nebes, R. D., & Oakley, M. A. (1991). A longitudinal study of word reading ability in Alzheimer's disease: Evidence from the National Adult Reading Test. *Cortex: A Journal Devoted to The Study of The Nervous System and Behavior, 27*(3), 367-376. doi:10.1016/S0010-9452(13)80032-9
- Goldstein, F. C., Gary, H. E. Jr., & Levin, H. S. (1986). Assessment of the accuracy of regression equations proposed for estimating premorbid intellectual functioning on the Wechsler adult intelligence scale. *Journal of Clinical and Experimental Neuropsychology, 8*(4), 405-412.
- Graf, P., & Uttl, B. (1995). Component processes of memory: Changes across the adult lifespan. *Swiss Journal of Psychology, 54*, 113-130.
- Hathaway, S. R., McKinley, J. C., & MMPI Restandardization Committee. (1989). *MMPI-2: Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring*. Ipswich, MA: University of Minnesota Press.

- Hawkins, K. A. (1995). Limitations to the validity of the Barona Regression Formula and similar demographically-based methods of estimating pre-injury intellectual functioning. *Behavioral Sciences & The Law*, 13(4), 491- 503.
doi:10.1002/bsl.2370130405
- Hedden, T., & Gabrieli, J. E. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience*, 5(2), 87-96.
doi:10.1038/nrn1323
- Hoofien, D., Vakil, E., & Gilboa, A. (2000). Criterion validation of premorbid intelligence estimation in persons with traumatic brain injury: 'Hold/dont hold' versus 'Best performance' procedures. *Journal of Clinical and Experimental Neuropsychology*, 22(3), 305-315. doi:10.1076/1380-3395(200006)22:3;1V;FT305
- Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26(2), 107-129. doi:10.1016/0001-6918(67)90011-X
- Johnstone, B., Callahan, C. D., Kapila, C. J., & Bouman, D. E. (1996). The comparability of the WRAT-R Reading Test and NAART as estimates of premorbid intelligence in neurologically impaired patients. *Archives of Clinical Neuropsychology*, 11(6), 513-519. doi:10.1016/0887-6177(96)82330-4
- Johnstone, B., Slaughter, J., Schopp, L., McAllister, J. A., Schwake, C., & Luebbering, A. (1997). Determining neuropsychological impairment using estimates of premorbid intelligence: Comparing methods based on level of education versus reading scores. *Archives of Clinical Neuropsychology*, 12, 591-601.

- Jones, G. N., Proto, D. A., Musso, M. W., Barker, A. A., & Gouvier, W. D. (2012). Using the NAART to estimate WRAT-4 reading grade level in a predominantly African American sample. *Applied Neuropsychology: Adult, 19*(4), 257-262.
doi:10.1080/09084282.2011.643966
- Kareken, D. A. (1997). Judgment pitfalls in estimating premorbid intellectual functioning. *Archives of Clinical Neuropsychology, 12*(8), 701-709.
- Kareken, D. A., Gur, R. C., & Saykin, A. J. (1995). Reading on the wide range achievement test revised and parental education as predictors of IQ: Comparison with the Barona formula. *Archives of Clinical Neuropsychology, 10*, 147-157.
- Kareken, D. A., & Williams, J. M. (1994). Human judgment and estimation of premorbid intellectual function. *Psychological Assessment, 6*, 83-91.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement: Second edition- comprehensive form*. Circle Pines, MN: American Guidance Service.
- Klesges, R. C., Wilkening, G. N., & Golden, C. J. (1981). Premorbid indices of intelligence: A review. *Clinical Neuropsychology, 3*(3), 32-39.
- Larrabee, G. J., Lergen, J. W., & Levin, H. S. (1985). Sensitivity of age-decline resistant ("Hold") WAIS subtests to Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 7*(5), 497-504.
- Law, R., & O'Carroll, R. E. (1998). A comparison of three measures of estimating premorbid intellectual level in dementia of the Alzheimer type. *International Journal of Geriatric Psychiatry, 13*(10), 727-730. doi:10.1002/(SICI)1099-1166(199810)13:10<727::AID-GPS851>3.0.CO;2-2

- Lezak, M. D., Howieson, D. B., & Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological Assessment* (4th ed.). New York, NY: Oxford University Press.
- Manly, J. J., Jacobs, D. M., Touradgji, P., Small, S. A., & Stern, Y. (2002). Reading level attenuates differences in neuropsychological test performance between African Americans and White elders. *The Journal of the International Neuropsychological Society*, 8, 341-348.
- Manly, J. J., Miller, S. W., Heaton, R. K., Byrd, D., Reilly, J., & Velasquez, R. J. (1998). The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. The HIV Neurobehavioral Research Center (HNRC) Group. *Journal International Neuropsychological Society*, 4, 291-302.
- Matarazzo, J. D., & Prifitera, A. (1989). Subtest scatter and premorbid intelligence: Lessons from the WAIS—R standardization sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1(3), 186-191. doi:10.1037/1040-3590.1.3.186
- McFie, J. (1975). *Assessment of organic intellectual impairment*. New York: Academic Press, Inc.
- Meehl, P. E. (1954). *Clinical and statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

- Mortensen, E. L., Gade, A., & Reinisch, J. M. (1991). A critical note on Lezak's best performance method in clinical neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, *13*, 361-371.
- Moss, A. R., & Dowd, T. (1991). Does the NART hold after head injury? A case report. *British Journal of Clinical Psychology*, *30*(2), 179-180.
doi:10.1111/j.20448260.1991.tb00934.x
- National Center for Education Statistics (2003). State and County Estimates of Low Literacy. Retrieved from:
<https://nces.ed.gov/naal/estimates/stateestimates.aspx#f2>
- Nelson, H. E. (1982). *National Adult Reading Test (NART): Test Manual*. NFER-Nelson: Windsor.
- Nelson, H. E., & McKenna, P. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology*, *14*(3), 259-267.
- Nelson, H. E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex: A Journal Devoted to The Study of the Nervous System and Behavior*, *14*(2), 234-244.
doi:10.1016/S0010-9452(78)80049-5
- Nelson, H. E., & Willison, J. (1982). *National Adult Reading Test, second edition*. London, England: GL Assessment.
- Paolo, A. M., & Ryan, J. J. (1992). Generalizability of two methods of estimating premorbid intelligence in the elderly. *Archives of Clinical Neuropsychology*, *7*, 135-143.

- The Psychological Corporation (2002). *Wechsler Individual Achievement Test: Second edition*. San Antonio, TX: Author.
- ProLiteracy (2016). U.S. adult literacy facts [PDF file]. Retrieved from https://proliteracy.org/Portals/0/pdf/PL_AdultLitFacts_US_flyer.pdf?ver=2016-05-06145137-067.
- Ricketts, J., Nation, K., & Bishop, D. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235-257.
- Russell, E. W. (1972). WAIS factor analysis with brain-damaged subjects using criterion measures. *Journal of Consulting and Clinical Psychology, 39*(1), 133-139.
- Ryan, J. J., Prifitera, A. (1990). The WAIS-R index for estimating premorbid intelligence: Accuracy of predicting short form IQ. *The International Journal of Clinical Neuropsychology, 12*, 20-23.
- Schaie, K. W. (2005). *Development influences on adult intelligence: The Seattle longitudinal study*. New York: Oxford University Press.
- Schinka, J. A., & Vanderploeg, R. D. (2000). Estimating premorbid level of functioning. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (pp. 39-67). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: Taylor & Francis Group.
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York, NY: Oxford University Press.

- Suhr, J. A. (2015). *Psychological assessment: a problem-solving approach*. New York: The Guilford Press.
- Sweet, J. J., Moberg, P. J., & Tovian, S. M. (1990). Evaluation of Wechsler adult intelligence scale- Revised premorbid IQ formulas in clinical populations. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(1), 41-44.
- Triola, M. F. (2014). *Elementary statistics using Excel*. Boston, MA: Pearson.
- Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1123-1137.
doi:10.1076/jcen.24.8.1123.8375
- Uttl, B., & Graf P. (1997). Color-word Stroop test performance across the adult life-span. *Journal of Clinical and Experimental Neuropsychology*, 19, 405-420.
- Wedding, D. (1983). Clinical and statistical prediction in neuropsychology. *Clinical Neuropsychology*, 5(2), 49-55.
- Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology*, 4(3), 233-265.
- Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.) Baltimore: Williams and Wilkins.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale- Revised manual*. New York: The Psychological Corporation.
- Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT 4: Wide range achievement test manual*. Lutz, FL: Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.

Yates, A. J. (1956). The use of vocabulary in the measurement of intellectual deterioration- A review. *The Journal of Medical Science*, 102, 409-438.