



Published in final edited form as:

J Clin Exp Neuropsychol. 2009 July ; 31(5): 605–610. doi:10.1080/13803390802375557.

Test–retest stability on the WRAT-3 reading subtest in geriatric cognitive evaluations

Lee Ashendorf^{1,2}, Angela L. Jefferson¹, Robert C. Green¹, and Robert A. Stern¹

¹Department of Neurology, Alzheimer's Disease Center, Boston University School of Medicine, Boston, MA, USA

²Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, USA

Abstract

The primary goal of this study was to establish the stability of the Wide Range Achievement Test (WRAT-3) Reading score across two annual assessments of aging individuals. Participants were classified as controls ($n = 200$), mild cognitive impairment (MCI; $n = 137$), or possible or probable Alzheimer's disease (AD; $n = 41$). Test–retest stability was acceptable to high for all diagnostic groups. The descriptive classification (e.g., “average”) remained consistent for only 74% of participants. Results indicated that WRAT-3 Reading scores are appropriate for use with older adults, though the use of categorical descriptors to describe premorbid ability based on these scores is not supported.

Keywords

Wide Range Achievement Test–Third Edition; Reading; Literacy; Test–retest; Geriatrics; Mild cognitive impairment; Alzheimer's disease

The Wide-Range Achievement Test–Third Edition (WRAT-3; Wilkinson, 1993) was developed for the assessment of scholastic achievement level in children and adults. The Reading subtest, due to its measurement of reading aloud irregularly spelled words, has been applied as an estimate of premorbid intelligence (Lezak, Howieson, & Loring, 2004). More recently, it has been used to estimate education quality, based on the utility of such reading tests in the assessment of literacy levels among multiethnic samples (Cosentino, Manly, & Mungas, 2007; Manly, Jacobs, Touradji, Small, & Stern, 2002). In fact, literacy level, as measured by WRAT-3 Reading performance, is a better predictor of memory decline than total years of education (Manly, Touradji, Tang, & Stern, 2003).

The utility of the WRAT-3 Reading test to estimate literacy or premorbid intelligence, though, is reliant on its stability across evaluations. According to the WRAT-3 manual (Wilkinson,

© 2008 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

Address correspondence to Robert A. Stern, Alzheimer's Disease Center, Boston University School of Medicine, 715 Albany Street, B7800, Boston, MA 02118-2526, USA (bobstern@bu.edu).

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

Publisher's Disclaimer: The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

1993), a standard error of measurement score equal to 5 is appropriate for use with either the blue or the tan WRAT-3 Reading subtest form. Similarly, research examining WRAT-3 Reading performance in adults across clinical and nonclinical samples indicates reasonably stable performance at varying test–retest intervals (McCaffrey, Duff, & Westervelt, 2000). One exception was Johnstone and Wilhelm’s (1996) finding that, for individuals with acquired cognitive dysfunction whose cognitive abilities improve, the mean WRAT-Revised (WRAT-R) Reading score also improved. These authors noted substantial variability in reading score over time, recommending cautious interpretation of the WRAT-R or WRAT-3 as an estimate of premorbid ability. Additionally, previous investigations of reading ability in individuals with dementia revealed decline (Storandt, Stone, & LaBarge, 1995; Taylor et al., 1996). Therefore, it is important to establish whether the WRAT-3 Reading subtest is a reliable measure for estimating the education quality or premorbid intellectual functioning of aging adults with normal cognition, preclinical dementia, and clinical dementia.

In the present study, the stability of WRAT-3 Reading subtest performance was examined by comparing two serial neuropsychological evaluations conducted one year apart. We hypothesized that individuals diagnosed with possible or probable Alzheimer’s disease (AD) would display greater variability in WRAT-3 Reading raw or scaled score than would cognitively normal adults and those with mild cognitive impairment (MCI).

METHOD

Participants

All study participants were enrolled in the Boston University Alzheimer’s Disease Core Center (BU-ADCC) patient/control registry, which longitudinally follows older adults with and without memory problems. Sample characteristics have been described in detail previously (Ashendorf et al., 2008; Jefferson et al., 2006, 2007). Briefly, inclusion criteria require that participants be community dwelling and English speaking, and that they have a study partner (to provide collateral information about functioning). Exclusion criteria include a history of major psychiatric illness (e.g., schizophrenia or bipolar disorder), other neurological illness (e.g., stroke or epilepsy), or head injury with loss of consciousness. The results of two consecutive annual evaluations (mean interval between evaluations = 13.9 months, $SD = 2.8$) were used for each participant in this retrospective study. Potential cases were only included if two consecutive evaluations were conducted and if the WRAT-3 Reading subtest was completed on both occasions.

Archival data were obtained from 378 participants (61% female, 83% White) with a mean age of 71.7 years ($SD = 8.7$) at the baseline evaluation. Participants included 200 individuals diagnosed as cognitively normal controls defined as having all cognitive performances within the normal range and a Clinical Dementia Rating (CDR; Morris, 1993) Global Score of 0.0. A total of 73 participants met Petersen (2004) criteria for MCI at the baseline evaluation and were labeled “probable MCI” based on a decline from previous level of functioning, lack of dependence in instrumental activities of daily living, cognitive complaint by the participant or study partner, and objective impairment in one or more cognitive domains (i.e., neuropsychological performance ≥ 1.5 standard deviations below normative data; Jefferson et al., 2007; Jefferson et al., 2008a, 2008b). An additional 64 participants met the above criteria but did not have a cognitive complaint by self or study partner and were classified as “possible” MCI. CDR Global Scores for this group were either 0.0 or 0.5. The AD sample included 41 participants meeting NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association) criteria for probable ($n = 21$) or possible ($n = 20$) AD (McKhann et al., 1984). CDR Total Scores for this group were either 1.0 or 2.0.

Procedure

The WRAT-3 Reading subtest was administered as part of a comprehensive, single-session neuropsychological test protocol. The local Institutional Review Board approved data collection efforts, and all participants provided written informed consent prior to testing.

Data analysis

Descriptive statistics or frequencies were calculated on all demographic variables (i.e., age, education, sex, and race). The one-year stability of the WRAT-3 Reading subtest raw scores (using Pearson correlations) was calculated in three forms: for the entire sample; separately for the three subgroups (i.e., control, MCI, AD); and for the three subgroups based on race. Reliable change indices (RCIs) that consider practice effects with 90% and 95% confidence intervals were calculated for the entire sample (Temkin, Heaton, Grant, & Dikmen, 1999). The RCI is a prediction measure that allows the clinician to distinguish a score that is statistically unchanged over time from a score that represents “true” change in performance, in either direction.

The WRAT-3 Reading raw scores from both visits were converted to scaled scores for each participant using normative data (Wilkinson, 1993). For individuals over age 74 years, the available norms for ages 65–74 years (i.e., the highest published age bracket) were used, as recommended by Manly and colleagues (2002). These scaled scores were used to assign range classifications (e.g., 90–109 = average, 110–119 = high average; Wilkinson, 1993). Participants whose WRAT-3 interval change resulted in assignment to different classification groups on their first and second visits were identified, and the frequency of conversion to a different scaled score range was compared with diagnostic conversion.

RESULTS

Sample characteristics

Baseline descriptive statistics are provided in Table 1 for the entire sample and each diagnostic subgroup. There was an association between age and education level, $r = -.138$, $p = .007$, such that older individuals tended to have fewer years of formal education, as we have previously reported (Jefferson et al., 2007). There was no sex effect on age. Non-Hispanic Caucasian participants (mean age = 72.3, $SD = 8.3$) tended to be older than African-American participants (mean age = 68.5, $SD = 9.8$), $t(81) = 2.9$, $p = .005$.

The age range of the control group was 55 to 102 years (mean = 71.6, $SD = 8.5$), and the mean Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) score was 29.3 ($SD = 1.0$). A total of 73 (37%) of these participants were men. At a one-year follow-up evaluation, 35 (18%) of these participants were diagnosed with MCI.

The combined MCI sample ($n = 137$) ranged in age from 49 to 101 years ($M = 70.0$, $SD = 8.3$), and 42% (57 participants) were men. The mean MMSE was 28.4 ($SD = 1.7$). At the one-year follow-up, only 5 of these participants (4%) had converted to AD, all of whom had met full Petersen (2004) criteria for MCI at the initial evaluation. An additional 34 individuals (25%) were reclassified as cognitively normal controls at the second evaluation, 22 of whom (65% >) had initially been classified as possible (versus probable) MCI.

The AD patients had a mean age of 77.8 years ($SD = 7.8$; age range = 60 to 101). A total of 44% were men. The mean MMSE score for this group was 23.4 ($SD = 3.1$). A total of 5 of these participants (12%) were identified as reverting to MCI at the annual follow-up visit.

The diagnostic groups differed with respect to race, $F(2, 375) = 3.8, p = .023$, as the MCI group included a higher proportion of African-American participants than did the other two groups. There was also an effect of age, $F(2, 375) = 13.8, p < .001$, as the AD group was older than the other two groups.

Hypothesis testing

As expected, the raw WRAT-3 Reading score differed by diagnosis, $F(2, 375) = 34.0, p < .001$, in the expected direction (i.e., control participants > MCI participants > AD participants). African-American participants performed more poorly than Caucasian participants as a group, $t(69) = 5.6, p < .001$; this discrepancy held across all three diagnostic groups (all $p < .01$). There was no difference in WRAT-3 Reading scores between men and women, $t(376) = -0.6, p = .549$.

The test–retest reliability of the WRAT-3 for the entire sample was .90 ($p < .001$). The stability for non-Hispanic Caucasian participants was .84 and for African-Americans was .96 (both $p < .001$). The reliability coefficients for control, MCI, and AD subgroups based on initial diagnosis were .81, .92, and .90, respectively (all $p < .001$; Table 2). Among those whose clinical diagnosis did not change at the second evaluation ($n = 299$), identical scores were found for 36 of 164 (22%) controls, 19 of 99 (19%) MCI participants, and 3 of 36 (8%) AD participants.

The 90% RCI confidence interval for WRAT-3 Reading raw scores among the control group fell between -3.84 to 2.84 , while the 95% confidence interval ranged -4.48 to 3.48 . As previously reported (Heaton et al., 2001), reliable change within samples that are known to have clinical areas of impairment should not be evaluated on the same basis as findings among cognitively normal individuals. Consistent with this point, using the 95% confidence interval, 4% of those consistently identified as controls demonstrated abnormal change, while 7% of MCI participants' performances and 19% of AD participants' performances changed. Therefore, separate RCIs were developed for each diagnostic group (Table 3). Race did not influence reliable change, as only 7% of Caucasians and 4% of African-Americans demonstrated significant change.

Of the 378 participants in the overall sample, 280 (74%) had the same WRAT-3 Reading scaled score category label (e.g., "Average") at both testing sessions. This rate of consistency over time indicates that this categorical classification method yielded a relatively low reliability (Cohen's kappa = .541). Half of those whose classification changed ($n = 49$; 13% of entire sample) declined in standard score (SS), while the remaining half ($n = 49$; 13% of sample) improved. One individual improved by two labels (average, SS = 108, to superior, SS = 120). Of those who changed, the mean change in SS was 5.1 points ($SD = 2.8$; range 0–12). The frequency of change did not differ between Caucasian (26%) and African-American (27%) participants.

Of the 41 participants whose consensus diagnosis progressed (i.e., control to MCI or MCI to AD), 12 (29%) changed WRAT-3 Reading subtest category labels. A total of 12 (32%) of the 38 participants who reverted in diagnosis (i.e., MCI to control or AD to MCI) received a different WRAT-3 Reading subtest category label at the second evaluation. Of the 299 participants whose clinical diagnosis remained the same for both assessments, 74 (25%) changed WRAT-3 Reading subtest category labels. Among this group were 35 control participants (21% of all controls), 26 MCI participants (26% of all MCI participants), and 13 AD participants (36% of all AD participants; Table 4). Though some change could be explained by obtaining consecutive scores close to a cutoff (e.g., changing from SS = 111 to SS = 109), 16% of individuals with initial WRAT-3 Reading scaled scores who were not near such a cutoff

still received a different label at the second visit (as opposed to 38% of those whose initial score was near a cutoff).

DISCUSSION

The WRAT-3 Reading subtest is often used in dementia evaluations as an estimate of literacy level or education quality, especially in multiethnic samples. The present study examined the test–retest stability of the raw scores and the descriptive labels (e.g., average, low average). The raw scores were found to have high stability among controls and patients and among Caucasian and African-American participants. However, the WRAT-3 descriptive classifications (e.g., average, high average) changed in 26% of the sample, with a higher rate among individuals with AD (i.e., 36%).

These findings confirm the test–retest stability of WRAT-3 Reading scores in dementia evaluations. Stability coefficients for all subgroups in the current study ranged from acceptable to high. Because the WRAT-3 Reading subtest has been recommended for use as a measure of literacy level among diverse populations (Manly et al., 2002), we also established the test–retest stability for African-American participants in particular. In fact, the stability coefficient was stronger among African-Americans than Caucasians in this study, though the reason for this seemingly anomalous finding was unclear. This stability supports the psychometric integrity of the WRAT-3 Reading subtest in serial dementia evaluations within racially diverse populations.

The other main study outcome was finding that, in spite of the test’s good one-year retest stability, only 74% of participants received the same descriptive classification for WRAT-3 Reading performance at both evaluations. One might expect small changes in scores near a “cutoff (for example, scores near $SS = 110$, the “boundary” between the average range and the high average range) to increase the chances that a score would exceed the cutoff at follow-up, but this explanation does not appear to account for the findings in this study. Rather, the results demonstrated that, of the participants whose descriptive labels did change at follow-up, those participants initially furthest from a cutoff were actually more susceptible to changes than those nearer to the cutoff. This effect reflects subtle but clinically noteworthy fluctuations in raw scores between annual visits.

The present findings indicate that caution should be used when assigning descriptive labels to an individual’s premorbid abilities or literacy level on the basis of WRAT-3 Reading scaled scores. While the general level of performance remains consistent (within 12 scaled score points from one assessment to the next for all participants), the probability of remaining within a descriptive category is lower than may be assumed by clinicians who use the WRAT-3 Reading subtest to describe a patient’s premorbid ability level.

The present study is not without limitations. The AD group is relatively small in comparison to the other groups, in large part due to the fact that many of the registry’s AD participants could not participate in formal testing. In addition, the groups differed on several demographic variables, including age ($AD > controls > MCI$) and race. The age difference may be explained by the fact that age is a strong risk factor for AD (Hebert et al., 1995). The higher proportion of African-American individuals in the MCI group may have occurred because diagnostic criteria for MCI are heavily based on neuropsychological test performances, which are known to differ between Caucasian and African American participants (Jefferson et al., 2007).

In summary, the test–retest stability of the WRAT-3 Reading subtest suggests that, across clinical and ethnic subgroups, there is little change over a one-year test–retest interval. The task therefore seems appropriate for use in deriving a numerical value against which to compare actual cognitive performance to determine level of impairment. In addition to addressing issues

pertaining to ethnic-minority performance in dementia evaluations, the WRAT-3 Reading subtest's use in studies of cognitive reserve is supported by the present results. However, the tendency to use categorical descriptors for an individual's particular premorbid baseline based on the WRAT-3 Reading subtest is not supported by these findings. As the new edition of the WRAT was recently published (WRAT-4; Wilkinson & Robertson, 2006), it will be important for future investigations to address whether the research conducted thus far using the WRAT-3 to estimate literacy levels, including the present study, will generalize to the WRAT-4.

Acknowledgments

This research was supported by P30-AG13846 (Boston University Alzheimer's Disease Core Center), M01-RR00533 (General Clinical Research Centers Program of the National Center for Research Resources, National Institutes of Health, NIH), R03-AG026610 (A.L.J.), R03-AG027480 (A.L.J.), K12-HD043444 (A.L.J.), and K23-AG030962 (Paul B. Beeson Career Development Award in Aging; A.L.J.).

REFERENCES

- Ashendorf L, Jefferson AL, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology* 2008;23:129–137. [PubMed: 18178372]
- 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. [PubMed: 17286880]
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;12:189–198. [PubMed: 1202204]
- Heaton RK, Temkin N, Dikmen S, Avitable N, Taylor MJ, Marcotte TD, et al. Detecting change: A comparison of three neuropsychological methods, using normal and clinical samples. *Archives of Clinical Neuropsychology* 2001;16:75–91. [PubMed: 14590193]
- Hebert LE, Scherr PA, Beckett LA, Albert MS, Pilgrim DM, Chown MJ, et al. Age-specific incidence of Alzheimer's disease in a community population. *Journal of the American Medical Association* 1995;273:1354–1359. [PubMed: 7715060]
- Jefferson AL, Byerly LK, Vanderhill S, Lambe S, Wong S, Ozonoff A, et al. Characterization of activities of daily living in individuals with mild cognitive impairment. *American Journal of Geriatric Psychiatry* 2008a;16:375–383. [PubMed: 18332397]
- Jefferson AL, Lambe S, Moser DJ, Byerly LK, Ozonoff A, Karlawish JT. Decisional capacity for research participation in individuals with mild cognitive impairment. *Journal of the American Geriatrics Society* 2008b;56:1236–1243. [PubMed: 18482298]
- Jefferson AL, Wong S, Bolen E, Ozonoff A, Green RC, Stern RA. Cognitive predictors of HVOT performance differ between individuals with mild cognitive impairment and normal controls. *Archives of Clinical Neuropsychology* 2006;21:405–412. [PubMed: 16893623]
- Jefferson AL, Wong S, Gracer TS, Ozonoff A, Green RC, Stern RA. Geriatric performance on an abbreviated version of the Boston Naming Test. *Applied Neuropsychology* 2007;14:215–223. [PubMed: 17848132]
- 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. [PubMed: 9375176]
- Lezak, MD.; Howieson, DB.; Loring, DW. *Neuropsychological assessment*. Vol. 4th ed.. New York: Oxford University Press; 2004.
- Manly JJ, Jacobs DM, Touradji 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. [PubMed: 11939693]
- Manly JJ, Touradji P, Tang M, Stern Y. Literacy and memory decline among ethnically diverse elders. *Journal of Clinical and Experimental Neuropsychology* 2003;25:680–690. [PubMed: 12815505]

- McCaffrey, R.J.; Duff, K.; Westervelt, HJ. Practitioner's guide to evaluating change with neuropsychological assessment instruments. New York: Plenum Press; 2000.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan E. Clinical diagnosis of Alzheimer's disease: Report of the NINCDSADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology* 1984;34:939–944. [PubMed: 6610841]
- Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* 1993;43:2412–2414. [PubMed: 8232972]
- Petersen R. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine* 2004;256:183–194. [PubMed: 15324362]
- Storandt M, Stone K, LaBarge E. Deficits in reading performance in very mild dementia of the Alzheimer type. *Neuropsychology* 1995;9:174–176.
- Taylor KI, Salmon DP, Rice VA, Bondi MW, Hill LR, Ernesto CR, et al. Longitudinal examination of American National Adult Reading Test (AMNART) performance in dementia of the Alzheimer type (DAT): Validation and correction based on degree of cognitive decline. *Journal of Clinical and Experimental Neuropsychology* 1996;18:883–891. [PubMed: 9157111]
- 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]
- Wilkinson, GS. *The Wide Range Achievement Test: Manual*. Vol. 3rd ed.. Wilmington, DE: Wide Range; 1993.
- Wilkinson, GS.; Robertson, GJ. *WRAT4: Wide Range Achievement Test professional manual*. Lutz, FL: Psychological Assessment Resources; 2006.

Initial demographic characteristics of each diagnostic subgroup and the entire sample at the first of two consecutive annual evaluations

TABLE 1

	N	Sex (% male)	Age (years)	Education (years)	Race (% AA)	MMSE
Control	200	37	71.6(8.5)	16.5 (2.9)	12	29.3 (1.0)
MCI	137	42	70.0 (8.3)	15.1 (3.3)	23	28.4(1.7)
AD	41	44	77.8 (7.8)	13.7 (3.2)	17	23.4(3.1)
Total	378	39	71.7(8.7)	15.7 (3.2)	17	28.3 (2.4)

Note. Standard deviations in parentheses. AA = African-American; MMSE = Mini Mental State Examination; MCI = mild cognitive impairment; AD = Alzheimer's disease.

TABLE 2

WRAT-3 Reading raw score stability across participant groups

	N	Time 1 (SD)	Time 2 (SD)	r*
Control	200	52.3 (3.1)	51.8 (3.4)	.81
MCI	137	49.1 (6.1)	49.1 (6.0)	.92
AD	41	46.7 (5.9)	45.9 (5.7)	.90
Caucasian	315	51.4 (3.9)	51.2 (4.1)	.84
African-American	63	46.1 (7.4)	45.3 (6.9)	.96
Total	378	50.6 (5.1)	50.2 (5.1)	.90

Note. Standard deviations in parentheses. WRAT-3 = Wide Range Achievement Test–Third Edition; Time 1 = WRAT-3 Reading subtest score at the first visit; Time 2 = WRAT-3 Reading subtest score at the second visit; *SD* = standard deviation; MCI = mild cognitive impairment; AD = Alzheimer’s disease. Diagnosis is based on the first of two consecutive, annual evaluations.

* All $p < .001$.

TABLE 3
WRAT-3 reliable change index values for each diagnostic group and ethnicity

	N	M_{diff}	SD_{diff}	95% RCI	90% RCI
Control	200	-0.50	2.03	-4.48, 3.48	-3.84, 2.84
MCI	137	-0.01	2.36	-4.64, 4.62	-3.89, 3.87
AD	41	-0.83	2.64	-6.00, 4.34	-5.17, 3.51
Caucasian	315	-0.27	2.24	-4.66, 4.12	-3.95, 3.41
African-American	63	-0.78	2.18	-5.05, 3.49	-4.37, 2.81

Note. RCI = reliable change index. WRAT-3 = Wide Range Achievement Test-Third Edition; M_{diff} = mean change in score between Time 1 and Time 2; SD_{diff} = standard deviation of the M_{diff} ; MCI = mild cognitive impairment; AD = Alzheimer's disease. Control, MCI, and AD groups refer to those who received the diagnosis at the first of two consecutive, annual consensus diagnostic team meetings.

TABLE 4

Frequency of change in WRAT-3 Reading descriptive label across two consecutive annual evaluations

	Increased	No change	Decreased
Control	17 (8.5)	154 (77.0)	29(14.5)
MCI	25 (18.2)	99 (72.3)	13 (9.5)
AD	7 (17.1)	27 (65.9)	7(17.1)
Total	49 (13.0)	280 (74.1)	49 (13.0)

Note. Percentages in parentheses. Descriptive label: e.g., average, high average. WRAT-3 = Wide Range Achievement Test–Third Edition; MCI = mild cognitive impairment; AD = Alzheimer’s disease. “Increased” refers to individuals whose increase in WRAT-3 Reading score from Visit 1 to Visit 2 raised their descriptive label (e.g., from average to high average). “Decreased” refers to those whose decrease in score corresponded to a lower descriptive label at the second evaluation (e.g., change from high average to average).