Published in final edited form as:

Exp Aging Res. 2010 October ; 36(4): 426–452. doi:10.1080/0361073X.2010.507427.

Race-Related Cognitive Test Bias in the ACTIVE Study: A MIMIC Model Approach

Adrienne T. Aiken Morgan,

Department of Clinical and Health Psychology, College of Public Health and Health Professions,
University of Florida, Gainesville, Florida, USA

Michael Marsiske,

Department of Clinical and Health Psychology, College of Public Health and Health Professions,
University of Florida, Gainesville, Florida, USA

Joseph Dzierzewski,

Department of Clinical and Health Psychology, College of Public Health and Health Professions,
University of Florida, Gainesville, Florida, USA

Richard N. Jones,

Hebrew Senior Life, Boston, Massachusetts, USA

Keith E. Whitfield,

Department of Psychology: Social and Health Sciences, Arts and Sciences and Trinity College,
Duke University, Durham, North Carolina, USA

Kathy E. Johnson, and

Department of Psychology, Indiana University Purdue University, Indianapolis, Indiana, USA

Mary K. Cresci

College of Nursing, Wayne State University, Detroit, Michigan, USA

Abstract

The present study investigated evidence for race-related test bias in cognitive measures used in the baseline assessment of the ACTIVE clinical trial. Test bias against African Americans has been documented in both cognitive aging and early lifespan studies. Despite significant mean performance differences, Multiple Indicators Multiple Causes (MIMIC) models suggested most differences were at the construct level. There was little evidence that specific measures put either group at particular advantage or disadvantage and little evidence of cognitive test bias in this sample. Small group differences in education, cognitive status, and health suggest positive selection may have attenuated possible biases.

The goal of the present investigation was to examine whether there was evidence of race-related bias in cognitive tests used in a large trial of cognitive interventions with older adults. While much work has been done to examine mean level differences in cognitive performance between racial and ethnic groups (e.g., Manly, Jacobs, Sano, Bell, Merchant, Small, & Stern, 1998; Manly, Jacobs, Touradji, Small, & Stern, 2002; Whitfield, Fillenbaum, Pieper, Albert, Berkman, Blazer, Rowe, & Seeman, 2000), the present investigation extends the current literature by examining the extent to which such differences represent general differences at the level of the cognitive constructs of interest, or whether they also represent test-specific

effects (i.e., such that some tests show specific race-group differences above and beyond general group differences at the latent level). Race in this study is operationalized as a comparison between African American and White older adults selected for participation in a cognitive clinical trial. The term African American in this study is used to characterize individuals of African, African American, and African Caribbean descent, while the term White is used to include persons of European descent.

Race Differences in Late Life Cognition

Evidence from a corpus of studies on the influence of race on late life cognition suggest mean level differences in performance on a variety of cognitive measures between African American and White older adults. Differences, usually in the form of higher performance for Whites, have been reported for tests of intelligence (e.g., Heaton, Ryan, Grant, & Matthews, 1996; Kaufman et al., 1988; Kush, Watkins, Ward, Ward, Canivez, & Worrell, 2001; Vincent, 1991) and for cognitive screening tests/batteries (e.g., Escobar, 1986; Fillenbaum et al., 1990; Inouye, Albert, Mohs, Sun, & Berkman, 1993; Whitfield et al., 2000; Zsembik & Peek, 2001; Manly et al. 1998; Manly et al. 2002; Patton, Duff, Schoenberg, Mold, Scott, Adams, 2003; Unverzagt, Hall, Torke, Rediger, Mercado, Gureje, Osuntokun, & Hendrie, 1996). This persistently lower performance by African Americans, which is present throughout the working life span (Avolio & Waldman, 1994), leads to earlier and more frequent cognitive impairment classifications and diagnoses of Alzheimer's disease and other dementias in African American elders (e.g., Carlson, Brandt, Carson, & Kawas, 1998; Inouye et al., 1993; Manly et al., 1998; Marcopulos, McLain, & Giuliano, 1997; Ripich, Carpenter, & Ziol, 1997; Unverzagt et al., 1996, Whitfield, 2002). In light of this disparity in cognitive performance and classification between African American and White elders, several researchers have noted that the influence of certain demographic and health factors are infrequently considered in many existing published studies exploring racial group differences in cognition and should be more deliberately investigated (Whitfield et al. 2000; Izquierdo-Rorrera & Waldstein, 2002; Zsembik & Peek, 2001).

Education is commonly invoked as a major explanation for cognitive test differences between racial groups. Studies have indicated that African American adults, on average, are likely to have attained less formal education than White adults (e.g., Harper & Alexander, 1990), and these differences exist over and above cohort differences in education attainment (Adams-Price, 1993). While most studies have considered education attainment in terms of number of years, years of education alone has been an inadequate explanation for group differences in cognitive performance; controlling for group differences in educational attainment has generally not explained the lower performance of African American elders on cognitive and neuropsychological measures (Manly et al., 1998; Manly & Jacobs, 2002). Several have argued that "years of education" does not have similar meaning across older racial groups in the United States (Jones, 2003). Furthermore, this education non-equivalence is also attributable to the historical effects of school segregation in the US prior to 1954, and associated factors like lower education expenditures, shorter school years, and higher student/teacher ratios that were experienced by African American students (Loewenstein, Arguelles, Arguelles, & Linn-Fuentes, 1994; Whitfield & Wiggins, 2003; Manly et al., 2002).

Health is another possible explanation for racial group differences in cognition. Health disparities in hypertension, diabetes, heart disease and stroke are present throughout much of adulthood and have been associated with poorer health outcomes and neuropsychological impairment (National Center for Health Statistics (NCHS), 1990; Berkman & Mullen, 1997; Whitfield, Weidner, Clark, & Anderson, 2002, Miles, 2002, Gurland, Wilder, Lantigua, Stern, Chen, Killeffer, & Mayeux, 1999). One particularly important health predictor of cognition appears to be physical function, which predicts poorer performance on tasks like Boston

Naming, MMSE and Digit Symbol (Whitfield, Baker-Thomas, & Heyward, 1999, Whitfield et al. 2000; Rosano, Simonsick, Harris, Kritchevsky, Brach, Visser, Yaffe, & Newman, 2005). Physical function may reflect both impaired performance factors (e.g., use of a pencil), as well as generally impaired health and vitality, serving as an index of underlying biological/physiological contributors to function (Marsiske, Klumb, & Baltes, 1997).

While sociodemographic factors like education, literacy, and health/physical function may be a partial explanation for race differences in late life cognition, biased tests are also a possible explanation. Test bias is defined as a property of a test which causes it to be unusually sensitive to group differences. When a test is biased, it consistently leads to particularly poorer performance in one group versus another. Much of the work that has been done to understand race-related test bias has been conducted in the young, particularly children and adolescents of school age. Research has been focused on the appropriate selection of psychological test instruments in school settings (Knauss, 2003), educational and achievement testing (Banks, 2006), and language tests and corresponding normative data that may be biased against bilingual children (Saenz & Huer, 2003). While the presence of race-related test bias has not always been found, it has been acknowledged that disproportionate representation of racial minorities in special education programs in U.S. school systems (Skiba et al., 2002) contributes to performance differences on tests.

Fewer studies have examined the influence of test bias in older populations, which motivates the present study. In one study exploring racial item bias in an older adult sample, findings suggested that there was racially related differential item functioning (DIF) for Black/African American and White participants on a modified Telephone Interview for Cognitive Status (TICS) (Jones, 2003). These differences included DIF on several TICS items (name objects, count backwards from 20, serial sevens subtraction, and name the president/vice-president). This DIF accounted for most of the mean cognitive performance group difference found, while background variables of low education in the Black/African American group and high income in the White group accounted for the remaining difference (Jones, 2003). In another study examining test bias of dementia screening instruments, findings indicated the Mini-Cog detected cognitive impairment in a racially diverse sample as well as or better than the Mini-Mental State Exam (MMSE) (Borson, Scanlan, Watanabe, Tu, and Lessig, 2005). Further, the Mini-Cog was found to be less biased by low education and low literacy (Borson et al., 2005). In a review paper examining DIF and item bias among cognitive measures used to assess the elderly, the authors concluded that many items on three dementia screening measures (Mini-Mental Status Examination, the Short Portable Mental Status Questionnaire, and the Mattis Dementia Rating Scale) tended to yield different levels of performance across education and racial groups (Teresi, Holmes, & Ramírez, 2002). Tests that were shorter, included easier items, and relied less on language/literacy skills and more on over-learned memory skills were less sensitive to differences among individuals of varying educational and racial backgrounds (Teresi et al., 2002). This is consistent with findings by Aiken-Morgan, Marsiske, & Whitfield (2008) and Manly et al. (1998, 2002). Nevertheless, more work is needed to examine DIF in other screening and neuropsychological tests (Teresi et al., 2002).

Importantly, studies of late-life race-related bias in cognitive tests have so far been conducted primarily with neuropsychological and cognitive screening instruments. There has been a relative absence of such investigations in the normal cognitive aging literature. In part, this may reflect the low representation of persons of color in the broader cognitive aging literature. Thus, one goal of the present investigation was to expand the study of cognitive test bias to include a broader set of measures. Furthermore, the focus in this paper was not on differential item functioning, but on differential test functioning (i.e., "test bias") to determine whether some cognitive tests show specific patterns of advantage/disadvantage for particular racial groups, after controlling for global performance groups differences. This approach is not

unrelated to the idea of “differential diagnostic evidence” (Mast et al., 2002), which suggests that some measures may be more sensitive to certain kind of group differences (e.g., vascular versus Alzheimer’s dementia) than others.

MIMIC Models

One way of evaluating test bias is through the use of the multiple indicators, multiple causes (MIMIC) models, an application of structural equation modeling (Muthen, 1989). The present manuscript used MIMIC modeling to investigate equivalence of ability distributions and to measure calibrations, discriminations, and differential measure functioning/bias in African American and White older adults in the ACTIVE trial.

The MIMIC approach investigates whether a given covariate or set of covariates has a unique and direct effect on a given measure of a construct. A schematic of this approach is shown in Figure 1. Figure 1 displays a construct, an exogenous covariate, and three measures of the construct. In the MIMIC model, factor loadings (relationship between the measures and their construct) are estimated. In addition, the covariate is allowed to have direct effects on both the latent construct, and the specific measures/indicators of that construct. If a given measure has no unique relationship with the covariate, then all of its’ relationship with the covariate will be fully mediated by the latent construct. On the other hand, if the relationship between the covariate and a measure is stronger than this purely indirect path would imply, then one would require an additional path between the covariate and the measure to capture this additional relationship. The requirement of this additional path would suggest bias; that is, the measure shows a significantly stronger relationship to the covariate than expected based on its’ saturation by the latent construct. Expressed differently, this would mean that the residual variance in the measure (variance not explained by the common construct factor) shows additional relationship to the covariate, above and beyond that covariate’s indirect influence on the measure via the latent construct. To concretize these ideas, if the covariate were “race,” the presence of an additional significant path from “race” to a measure would suggest that that measure is more strongly related to “race” (i.e., evinces stronger race differences) than the construct to which it belongs.

The same logic applies in the opposite direction; a significant path would also result when a given measure is less related to the covariate (e.g., race) than the indirect relationship via the construct would imply. The MIMIC model will be considered again, more explicitly, when the specific model to be evaluated in this study is discussed in the Methods section of this paper.

It should be noted that test bias can be discussed in different ways. One way, the approach used in the current study, explores threats to the validity of a test that can occur in the scoring of individuals on underlying traits (this paper). A second approach to studying bias refers to the differential prediction of outcomes (i.e., in the use and interpretation of a test’s trait estimate), but that is not the focus of the current paper. The current study focused on bias as a concept internal to the trait.

A few prior studies have used MIMIC to examine differential item functioning and test bias in research and clinical samples. As mentioned above, Jones (2003) used MIMIC modeling to demonstrate DIF for a Black/African American sample on a modified TICS measure. Additionally, Mast and Lichtenberg (2000) used MIMIC to examine the effect of population heterogeneity, but not race, among geriatric patients on factor structure and DIF on a measure of functional independence (Functional Independence Measure; FIM). Their findings suggested three motor functioning and three cognitive functioning items showed DIF systematically across sample subgroups (young-old vs. old-old, male vs. female, and depressed vs. non-depressed), and the authors concluded that scores on the FIM may lead to biased comparisons of functional abilities across subgroups (Mast & Lichtenberg, 2000).

Finally, Mast, MacNeill, and Lichtenberg (2002) demonstrated the MIMIC methodology by integrating neuropsychological test data to a MIMIC model to examine the influence of cerebrovascular disease (CVD) on global cognitive impairment dimensions and individual tests, after controlling for cognitive impairment. Their findings demonstrated the presence of CVD in dementia was unrelated either to dimensions of global cognitive impairment or to performance on nine of ten neuropsychological tests (Mast et al., 2002). While Mast and colleagues did not consider race, they demonstrated the practical utility of applying the MIMIC approach to investigate possible sources of bias in commonly used gerontological and geriatric measures. The current study examines race-related cognitive differences, at both the latent and test-specific level, using MIMIC (Muthen, 1989) modeling.

Specific Aims

The present investigation extends the current literature by addressing two specific aims. First, we explored differences in mean cognitive test performance between African American and White community-dwelling older adults. Second, we used MIMIC modeling to specifically investigate racial group differences and possible test bias.

METHODS

The ACTIVE Study

The ACTIVE study is a randomized, controlled, single-masked clinical trial designed to examine whether cognitive training can affect cognitively based measures of daily functioning (Ball et al., 2002; Jobe et al., 2001). Data for the current study are taken from the ACTIVE baseline assessment, which was completed prior to the beginning of the cognitive intervention part of the trial. A variety of recruitment strategies (e.g., through local churches and senior organizations) were employed by each of six field sites (Birmingham, AL; Boston, MA; Indianapolis, IN; Baltimore, MD; Detroit, MI; and north central Pennsylvania), and are documented in Jobe et al. (2001).

Participants

There were a total of 2,802 participants in the sample; 61 of these older adults were excluded from the present analyses because either they did not identify themselves as African American/black or European American/white or were incorrectly randomized, and thus, represented categories of insufficient frequency to permit group comparisons. The present analyses were based on the remaining sub-sample of 2,741 adults aged 65 and older, of whom 75.8% were women, with a mean age of 73.6 years and a mean of 13.5 years of education (Table 1). The sample was comprised of approximately 26.6 percent ($n = 729$) African Americans. The African American group was significantly younger ($p < .001$) than the White group (*mean age* = 72.1 years vs. 74.1 years, respectively) and included significantly more females ($p < .001$; 83.2% versus 73.2% respectively). African Americans also reported having significantly fewer years of education ($p < .001$; *mean education* = 13.0 vs. 13.7 years, respectively) and had poorer self-rated physical function on the SF-36 (Ware & Sherbourne, 1992) ($p < .001$; *mean* = 65.0 vs. 70.2). For the SF-36 scale, we compared groups on all subscales and found the physical function subscale to be the only physical health subscale to differ significantly by race. Additionally, the SF-36 physical function subscale was the most related to cognition in this sample.

Exclusion Criteria—Potential participants in the ACTIVE study were screened to exclude individuals based on criteria related to seven factors: 1) age < 65 years at initial screening; 2) probable dementia, as defined by a score of 22¹ or lower on the Mini-Mental Status Examination (Folstein, Folstein, & McHugh 1975), a widely used screening tool, and/or a self-reported diagnosis of Alzheimer's disease; 3) substantial functional decline, as determined by

self-reported need for assistance in performing activities of daily living (ADL) related to bathing, dressing, or personal hygiene; 4) specific medical conditions, such as stroke, in the past 12 months, certain high-fatality cancers (e.g., liver, lung, esophagus), or current chemotherapy and or radiation treatment for any cancer; 5) severe sensory loss; 6) communication difficulties; and 7) recent or current participation in cognitive training studies; or unavailability during any phase of the study. The sample was not restricted to people born in the US, nor was it restricted to persons for whom English was the first language. Trained telephone screeners were asked to score whether participants could understand and make themselves understood, as part of the screening process. The rationale for these exclusion criteria related to the goals of overall clinical trial study is reviewed by Jobe et al. (2001).

Measures and Data Collection Procedures

Demographic data were collected via telephone interview. Cognitive data were collected over the course of two 90-minute testing sessions conducted individually and a third 120-minute session conducted in a small group. Age, gender, years of education, and physical function were used as covariates in all analyses to control for non-equivalent sampling by race (i.e., African Americans were younger, more likely to be female, and had poorer physical function) (Table 1). Insufficient data were available at the ACTIVE baseline assessment to permit us to include other covariates, such as income and socioeconomic status.

The cognitive battery consisted of measures of reasoning, speed, and memory. The chosen measures represent commonly used cognitive instruments to test cognitive performance in aging populations. In addition, the measures were selected to reflect the endpoints of the larger ACTIVE clinical trial. Baseline assessments of memory, reasoning, and speed of processing, which were the planned targets of training for the full study, were included in the present study, as delineated in Table 2. Variables are coded such that higher scores mean better performance for reasoning and memory measures, while lower scores on speed measures indicate better performance.

RESULTS

Analysis Plan

All structural equation models in this study were estimated using AMOS 16.0 program (Arbuckle, 2007). Structural equation models employed Blom-transformed cognitive scores, producing more normally distributed measures (Ball et al., 2002; Blom, 1958). Models were evaluated using overall fit indexes representing the root mean square error of approximation (RMSEA), a fit index indicating the discrepancy between the original and reproduced covariance matrix divided by the degrees of freedom and a fit index for which values of .08 or lower are indicative of adequate fit. Comparative fit index (CFI), incremental fit index (IFI) and the normed fit index (NFI) were also examined (Bentler, 1989; Bentler & Bonet, 1980; Bollen, 1989; Marsh, Balla, & McDonald, 1988), and were thought to be indicative of adequate fit when they assumed values of 0.95 or higher. Chi-square statistics were not examined as indicators of specific model fits, because of known problems of inflation in larger samples and with deviations from normality in variables under study (Akaike, 1987; Carmines & McIver,

¹The criterion of MMSE score ≥ 23 was used to ensure adequate recruitment of African American older adults to meet the goals of the pilot and overall ACTIVE clinical trial. It should be noted that MMSE cutoff was selected because it was lower than some published recommended cutoff scores for the MMSE (Crum, Anthony, Bassett, & Folstein, 1993), and was thought to be more broadly inclusive. In addition, the “one-size-fits-all” cutoff (Cullen et al, 2005) was developed to permit quick implementation during the initial screening phase of the study (i.e., testers quickly used it to determine initial eligibility in the field, so a more nuanced set of cutoffs divided by age, education and/or race was not feasible). The consequence of this cutoff did lead to significant race differences in exclusion due to the MMSE. Of the 316 Whites who were excluded, 115 (36.4%) received an MMSE below 23; of the 222 blacks who were excluded, 150 (67.6%) were excluded on the basis of an MMSE below 23. This represented a significant difference (with continuity correction, $\chi^2[N=538, df=1] = 49.46, p < .001$).

1981). At the same time, to compare nested models, a chi-square difference test was conducted. For nested models, the difference between the model chi-square statistics is distributed as a chi-squared statistic, with the difference in degrees of freedom as the *df* of the difference statistic. In addition, Akaike's Information Criterion (AIC) was examined; better models should evince smaller AIC values.

Mean Cognitive Performance Differences

To explore differences in mean cognitive test performance between community-dwelling African American and White older adults group means were compared, using *t*-test analyses. These comparisons were done to determine whether significant differences in performance existed between race groups on the selected cognitive measures, controlling for demographic variables and data collection site. As indicated in Table 3, there were significant mean differences between the African American and White groups, reflecting better performance for Whites, for all cognitive measures. For several measures, there was non-homogeneity of variance, thus *t*-values and degrees of freedom for these *t*-test comparisons were adjusted for unequal variances.

Ability Factor Model and Invariance

Before testing the MIMIC model, including race group differences at both the level of the cognitive factors and specific tests, the general fit of the proposed three-factor model of cognition (reasoning, memory, speed of processing) was examined. A confirmatory factor analysis, using AMOS 16.0, was conducted. The planned MIMIC model required a combined model across both race groups, so that a dummy variable representing race could be included as a predictor in the model. However, to evaluate the feasibility of combining the data from both racial groups, we first estimated our cognitive factor model while examining factor invariance across the two groups (models were estimated using covariance metric). Thus, the three-factor cognitive ability structure was estimated simultaneously in the African American and White samples, and the fit in both groups was evaluated. Invariance tests examined the effects on fit of three different levels of invariance constraint (see Table 4). For all models tested, uniqueness terms (variance in each measure not explained by its factor) were allowed to vary between groups in all models. A partially invariant solution, with loading invariance between groups, represented the best model and also had the lowest levels of AIC.

Completely standardized factor loadings for the combined groups have been included in Table 5. In the White sample, correlations between the cognitive factors were as follows: reasoning and memory $r = 0.66$; reasoning and speed, $r = -0.66$; memory and speed, $r = -0.58$. In the African American sample, correlations between the cognitive factors were as follows: reasoning and memory $r = 0.63$; reasoning and speed, $r = -0.53$; memory and speed, $r = -0.47$.

MIMIC Model Results

Next, MIMIC modeling examined whether there was specific race-related test bias in the ACTIVE baseline cognitive battery. MIMIC models allowed for the influence of multiple factors (e.g., race, age, gender, education, physical function) to be considered when determining whether measures of cognitive factors function similarly cross-culturally. An extension of confirmatory factor analysis, a MIMIC model is a structural equation model (SEM) that has one or more latent variables simultaneously identified by multiple endogenous indicators and multiple exogenous causal variables (Muthen, 1989). Endogenous indicators in this paper are the tests that compose a latent cognitive factor, while exogenous indicators are demographic variables of race, age, gender, education, and physical function.

A schematic of the estimated 3-factor MIMIC model is shown in Figure 2. In this figure, paths are shown only for one latent cognitive factor for simplicity. Paths from each latent cognitive

variable (reasoning, memory, and speed) to its corresponding cognitive tests (path *e*) reflect factor loadings. MIMIC models were run in two steps. In the first step, paths were estimated to the latent construct from race (path *a*), while simultaneously controlling for exogenous demographic covariates (age, gender, education, physical function, as well as testing site [not shown]; path *b*). In a second model, paths to the latent constructs were fixed (to the values obtained in the preceding model), and paths were estimated from race to each specific test (path *c*), controlling for the effects of the demographic covariates on that test (path *d*). The models were run in these two steps to permit identification, which is akin to procedures used in extension analysis (Kemper & Summer, 2001; Schaie et al., 2005). Note that this model employs a combined model, collapsed across race groups. This combined model was justified based on 1) earlier findings of two-group factor invariance in loadings, 2) the fact that a highly constrained model with both factor loadings and covariances forced to equality also fit well, and 3) the technical requirement of MIMIC models to have a combined solution; most MIMIC model studies do not first test invariance (Muthen, 1989). It should be noted that for dichotomous variables, like race, the path coefficient expresses the mean difference between groups; that is, what is the unit difference in the dependent variable (i.e., specific measure, construct) associated with the difference between groups.

The logic of the MIMIC test is as follows. Race differences on a given test (e.g., Letter Series) may be due to two sources. One source is because the underlying construct which produces performance on the Letter Series test (i.e., Reasoning) itself evinces race differences. Thus, race differences at the construct level reflect themselves in race differences on the specific measures which represent that construct. However, if there is additional variance in the Letter Series test which is unique to that measure (i.e., not reflective of the underlying common factor; in other words, the “unique” or “residual” variance in Letter Series), and if that unique variance in Letter Series is *also* related to race, then there may be an *additional, direct* relationship between race and the Letter Series. What this would mean is that the Letter Series test evinces race differences for two reasons: because (a) reasoning shows race differences, and (b) the specific test contains further race differences above-and-beyond the reasoning construct. This additional race-difference variance in the Letter Series is the *bias*. It says that Letter Series shows more race differences than we would expect on the basis on the Reasoning construct alone.

Each cognitive construct was modeled in two steps. The first step, which estimated regression coefficients between race and the cognitive factors, controlling for covariates (race, age, gender, education, and physical function) and data collection site is summarized in Table 6; coefficients shown are standardized regression weights. All regression coefficients (ranging between .05 and .50 in magnitude) were statistically significant, except for the coefficients between gender and the speed ($b = .00$) factor. Of special interest were the regression coefficients for the effect of race on the latent cognitive factors, which were statistically significant and of low to moderate magnitude. Standardized b -weights were: reasoning = $-.36$, memory = $-.26$, and speed = $.15$. The fit of these models were acceptable: Reasoning $\chi^2(20) = 64.46$, $p < .001$; NFI = 0.99; IFI = 0.99; RMSEA = 0.028; CFI = 0.99; Model AIC = 232.46; Memory $\chi^2(20) = 309.81$, $p < .001$; NFI = 0.96; IFI = 0.96; RMSEA = 0.073; CFI = 0.96; Model AIC = 477.81; Speed $\chi^2(32) = 376.76$, $p < .001$; NFI = 0.95; IFI = 0.96; RMSEA = 0.063; CFI = 0.96; Model AIC = 550.76.

Next, in the second step, the existence of specific residual paths between race and the individual tests, after controlling for covariates, site, and the effects of the predictors on the latent construct, was examined. Positive coefficients suggested race group differences at the test level were less than expected based on group differences at the factor level. Table 7 displays the standardized regression coefficients obtained from these models. There were small but significant residual performance differences for African American participants on two

measures: the Rivermead Behavioral Memory scale (Paragraph Recall) (-0.05), and UFOV Subtest 2 (0.05) ($p < .05$; Table 7). For both tests, these paths indicated tiny test-specific performance disadvantages for African Americans. That is, the magnitude of the residual race difference was more than that of the overall latent level difference attributable to race. The fit of these models were acceptable: Reasoning $\chi^2(2) = 0.90$, $p > .05$; NFI = 1.00; IFI = 1.00; RMSEA = 0.000; CFI = 1.00; Model AIC = 204.90; Memory $\chi^2(2) = 1.72$, $p > .05$; NFI = 1.00; IFI = 1.00; RMSEA = 0.000; CFI = 1.00; Model AIC = 205.72; Speed $\chi^2(5) = 220.40$, $p < .001$; NFI = 0.97; IFI = 0.97; RMSEA = 0.125; CFI = 0.97; Model AIC = 448.40. Because these second steps were not nested within the ones above (the second steps fixed elements of the preceding solutions, so the models could be identified), hierarchical model tests could not be conducted. Nonetheless, these results indicated that above the expected race group performance difference at the level of the latent constructs of reasoning, memory, and speed, there were also test-specific differences that showed a slight consistent pattern of disadvantage for African American elders in this sample. These significant effects were nonetheless very weak (Cohen, 1988), and their significance may in part be attributable to the substantial statistical power afforded by the relatively large size of the multi-site sample.

DISCUSSION

The goal of the present study was to examine the presence of race-related test bias in the ACTIVE study sample of African American and White older adults. Test bias in the present context was defined as a property of a test which caused it to be unusually sensitive to group differences and consistently lead to poorer performance in one racial group versus another. While various sources of racial group differences have been examined in previous research (e.g., education, literacy, physical function), little work to date has examined the role of test-bias in understanding group differences. Results of this study found typical racial group mean differences on various measures of reasoning, speed, and memory. Further, racial group differences were also found at the latent construct level. Regarding cognitive factor invariance, analyses supported strict invariance of factor loadings across groups.

Using MIMIC modeling, we found little evidence of test-specific bias; for most measures, residual effect of race on specific tests was near-zero, and only reached significance in two cases. In both cases (1 involving memory and 1 for speed), the residual effect of race was less than 0.10 and likely reached significance only due to the substantial statistical power of this study. However, the pattern of race effects was consistent; the residual effect of African American status was positive.

Thus, there are several conclusions that can be made based on these findings. First, as has been documented consistently, racial differences in cognitive performance are clearly present in late life. Next, unlike early in the life span, or even other studies of late-life mental status, there was little evidence of test-specific bias here. This suggests that over and above important potential mediating factors (i.e., education, literacy, physical function, and socioeconomic status) that have been found to explain cognitive performance differences between African Americans and Whites in previous work, individual cognitive tests added little to no added bias for African Americans, which is an encouraging and positive finding.

Nevertheless, this study has several limitations that are important to consider. The ACTIVE study was not designed explicitly for the purpose of cognitive comparison between African American and White older adults. In addition, using a selective convenience sample means that generalizability to the broader population of older adults requires further consideration. While essentially a sample of convenience, our sample of African Americans and Whites were reasonably well matched on mean level of education. In the US, it is almost certain that racial discrimination made access to education highly discrepant across African American and White

groups (Whitfield & Wiggins, 2003). The African Americans in this study, who on average completed a thirteen years of education, are less typical overall of African American elders. According to the 2000 US Census, for the cohort of adults aged 65 and older in 2000, the median education for African Americans was between 9 and 11 years. In contrast, the Whites in the present sample averaged 13.7 years of school, which is more typical of the median education of White Americans in this cohort (US Census, 2002). We therefore acknowledge the possibility that the selection process and the level of mean education for this study lessened the likelihood of finding cognitive test performance bias in this sample.

We expect our findings would have yielded greater evidence of test bias had we included more individuals with lower levels of education attainment, specifically in the African American group. Further, it is a concern that bias might have been larger in a sample that better represented individuals of lower education. This is a potential limitation and requires follow-up study in a more educationally heterogeneous sample to see whether the cognitive test bias varies by educational level. Nonetheless, the present study could be taken as a “natural experiment” with a sample of positively selected African Americans in terms of educational attainment.

There remains the need to understand and explain why persistent mean and latent level racial group differences remain. These differences have implications for late life impairment rates as well as implications for the long-term effects of early intervention attempts to reduce educational and health disparities. This assertion may extend to the cognitive status screening, given that only participants with MMSE scores of 23 or higher were admitted into the study. Some investigators have argued that different MMSE cutoffs should be used for African American and White elders (Bohnstedt, Fox, & Kohatsu, 1994). The decision of the ACTIVE investigators not to implement differential cut-offs, which was done both to address field logistics (participants needed to be screened in “on the spot”) and to ensure a common minimum performance baseline in all participants who might later be randomized to a training condition, potentially might have further have led to disproportionately positive selection bias for the African American group. On the other hand, from the perspective of quasi-experimental methodology, the sample selection resulted in two race groups with fewer gross differences than in the population, permitting a more careful evaluation of the unique effect of race, if any, on test bias.

Speculatively, the MMSE cutoff of 23 essentially trimmed the lower portion of the cognitive status distribution. Moreover, given the aforementioned known race differences in the MMSE distribution, it is likely that this cutoff disproportionately removed African Americans. Indeed, a total of 3,357 ACTIVE African American and White participants were screened with the MMSE. Of these, 95% of White participants (2,279) and 84% of African American participants (811) were MMSE eligible. This difference in eligibility was significant [$\chi^2(1) = 107.78, p < .0001, \theta = 0.18$]. Beyond the obvious generalizability limitations that this finding implies, there is no way (absent a new data collection) to determine whether bias findings would have been larger if the lower-MMSE groups could have been included. At the same time, there is little extant empirical reason to believe that bias, if it exists, only exists in the “lower portion” of the cognitive ability distribution. For now, however, the present findings seem to permit the assertion, with some statistical power, that there is little evidence of race-related bias in the top 85% (i.e., cognitively unimpaired segment) of the aging population. Nonetheless, the constrained generalizability of the present sample is again acknowledged as a limitation, even as the sample is more heterogeneous than most in the cognitive aging literature.

Adequate data on socioeconomic status (SES) were not available for the ACTIVE baseline assessment to permit analysis of the possible influences of this variable; thus, SES could not be controlled for in this study. The present study therefore cannot determine the extent to which our findings were influenced by uncontrolled SES effects. Additionally, the data were collected

to measure fluid abilities (e.g., information processing), rather than crystallized abilities that are more closely related to education and acculturation and found to be more biased in assessment. Measures of fluid ability are thought to be more related to biological changes that occur with age, which might be reflected in health differentials. Thus, it is not known whether inclusion of measures of crystallized knowledge in this sample would have resulted in greater evidence for test bias.

In sum, these results suggest few reasons not to use common cognitive measures with African American community-dwelling elders. Nevertheless, additional research is essential for determining whether test bias presents more of a problem when older adults are in poor health, have very little education, or are socioeconomically vulnerable. It is therefore necessary to examine a broader selection of cognitive instruments and in a sample that is more diverse and more representative of the lower end of the health, education, and SES spectrums. Group differences then must arise from other covariates not included in this analysis. Individual factors such as literacy and health status probably account for the lower mean scores found for African Americans (Manly et al. 1998, 2002; Whitfield, 1996; Whitfield & Willis 1998). The inclusion of other covariates is needed to fully comprehend the origin and source of group differences in cognitive aging. Finally, while it was not the focus of the current study, results of the MIMIC models demonstrated significant test bias against women on the Letter Sets test. Thus, future inquiry should also explore other potential sources of cognitive test bias in older adults.

Acknowledgments

This study was funded by grant U01AG014276, including a Minority Supplement award, from the National Institute on Aging. We also thank the ACTIVE Steering Committee for permission to use the data, and for feedback on the initial proposal for this work (Karlene Ball, University of Alabama-Birmingham; Jonathan King, National Institute on Aging; Paul Cotton, National Institute of Nursing Research; George Rebok, Johns Hopkins University; Fred Unverzagt, Indiana University School of Medicine; Sherry L. Willis, The University of Washington)

References

- Adams-Price CE. Age, education, and literacy skills of adult Mississippians. *The Gerontologist* 1993;33:741–746. [PubMed: 8314100]
- Aiken Morgan AT, Marsiske M, Whitfield KE. Characterizing and explaining differences in cognitive test performance between African American and European American older adults. *Experimental Aging Research* 2008;34:80–100. [PubMed: 18189169]
- Akaike H. Factor analysis and AIC. *Psychometrika* 1987;52:317–332.
- Arbuckle, JL. *Amos 16.0 User's Guide*. SPSS, Inc; Chicago, IL: 2007.
- Avolio BJ, Waldman DA. Variations in cognitive, perceptual, and psychomotor abilities across the working life span: Examining the effects of race, sex, experience, education, and occupational type. *Psychology and Aging* 1994;9(3):430–442. [PubMed: 7999327]
- Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, Morris JN, Rebok GW, Smith DM, Tennstedt SL, Unverzagt FW, Willis SL. Effects of cognitive training interventions with older adults. *Journal of the American Medical Association* 2002;288:2271–2281. [PubMed: 12425704]
- Ball K, Owsley C. The useful field of view test: a new technique for evaluating age-related declines in visual function. *Journal of the American Optometric Association* 1993;64:71–79. [PubMed: 8454831]
- Banks K. A comprehensive framework for evaluating hypotheses about cultural bias in educational testing. *Applied Measurement in Education* 2006;19(2):115–132.
- Benedict RHB, Schretlen D, Groninger L, Brandt J. Hopkins Verbal Learning Test – Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist* 1998;12(1): 43–55.
- Bentler, P. *EQS: Structural equations program manual*. BMDP Statistical Software Inc; Los Angeles, CA: 1989.

- Bentler P, Bonet D. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 1980;88:588–606.
- Berkman, LF.; Mullen, JM. How health behaviors and the social environment contribute to health differences between black and European American older Americans. In: Martin, LG.; Soldo, BJ., editors. *Racial Differences in the Health of Older Americans*. Washington, D.C: National Academy Press; 1997. p. 163–182.
- Blom, G. *Statistical Elements and Transformed Beta Variables*. Wiley; New York: 1958.
- Bohnstedt M, Fox P, Kohatsu N. Correlates of Mini-Mental Status Examination Scores among elderly demented patients: The influence of race/ethnicity. *Journal of Clinical Epidemiology* 1994;47(12): 1381–1387. [PubMed: 7730847]
- Bollen, K. *Structural equations with latent variables*. Wiley; NY: 1989.
- Borson S, Scanlan JM, Watanabe J, Tu S-P, Lessig M. Simplifying Detection of Cognitive Impairment: Comparison of the Mini-Cog and Mini-Mental State Examination in a Multiethnic Sample. *Journal of the American Geriatrics Society* 2005;53(5):871–874. [PubMed: 15877567]
- Carlson MC, Brandt J, Carson KA, Kawas CH. Lack of relation between race and cognitive test performance in Alzheimer's disease. *Neurology* 1998;50:1499–1501. [PubMed: 9596021]
- Carmines, EG.; McIver, JP. *Analyzing Models with Unobserved Variables: Analysis of Covariance Structures*. In: Bohnstedt, GW.; Borgatta, EF., editors. *Social Measurement: Current Issues*. Beverly Hills, CA: Sage Publications; 1981. p. 65–115.
- Crum RM, Anthony JC, Bassett SS, Folstein MF. Population-based norms for the Mini-Mental State Examination by age and education level. *Journal of the American Medical Association* 1993;269:2386–2391. [PubMed: 8479064]
- Cullen B, Fahy S, Cunningham CJ, Coen RF, Bruce I, Greene E, Coakley D, Walsh JB, Lawlor BA. Screening for dementia in an Irish community sample using MMSE: a comparison of norm-adjusted versus fixed cut-points. *International Journal of Geriatric Psychiatry* 2005;20:371–376. [PubMed: 15799072]
- Escobar JI. Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity: Cultural and linguistic artifacts. *Journal of Nervous and Mental Disease* 1986;174(10): 607–614. [PubMed: 3760851]
- Ekstrom, RB.; French, JW.; Harman, H.; Derman, D. *Kit of Factor Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service; 1976. Revised ed
- Fillenbaum GG, Heyman A, Williams K, Prosnitz B, Burchett B. Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly Black and White community residents. *Journal of Clinical Epidemiology* 1990;43:650–660.
- 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]
- Gurland BJ, Wilder DE, Lantigua R, Stern Y, Chen J, Killeffer EHP, Mayeux R. Rates of dementia in three ethnic groups. *International Journal of Geriatric Psychiatry* 1999;14(6):481–493. [PubMed: 10398359]
- Harper, MS.; Alexander, CD. Profile of the black elderly. In: Harper, MS., editor. *Minority aging: essential curricula content for selected health and allied health professions*. 1990. p. 193–222. DHHS Publication No. HRSR-DV 90–4
- Heaton, RK.; Ryan, L.; Grant, I.; Matthews, CG. Demographic influences on neuropsychological test performance. In: Grant, I.; Adams, KM., editors. *Neuropsychological assessment of neuropsychiatric disorders*. Vol. 2. New York: Oxford University Press; 1996. p. 141–163.
- Inouye SK, Albert MS, Mohs R, Sun K, Berkman JF. Cognitive performance in a high-functioning community-dwelling elderly population. *Journals of Gerontology Series A, Biological Sciences and Medical Sciences* 1993;48:146–151.
- Izquierdo-Porrera AM, Waldstein SR. Cardiovascular risk factors and cognitive function in African Americans. *Journals of Gerontology: Series B: Psychological Sciences & Social Sciences* 2002;57B(4):P377–P380.
- Jobe JB, Smith DM, Ball K, et al. ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials* 2001;4:453–479. [PubMed: 11514044]

- Jones RN. Racial bias in the assessment of cognitive functioning of older adults. *Aging & Mental Health* 2003;7(2):83–102. [PubMed: 12745387]
- Joreskog, KG.; Sorbom, D. LISREL8: Structural equation modeling with the SIMPLIS command language. Hillsdale, NJ: Erlbaum; 1993.
- Kaufman AS, McLean JE, Reynolds CR. Sex, race, residence, region and educational differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology* 1988;44(2):231–248. [PubMed: 3360941]
- Kemper S, Sumner A. The structure of verbal abilities in young and older adults. *Psychology and Aging* 2001;16(2):312–322. [PubMed: 11405318]
- Knauss LK. Research on psychological testing supervision using the Holloway (1995) method: A critique of DeCato (2002). *Psychological Reports* 2003;92(1):141–142. [PubMed: 12674272]
- Kush JC, Watkins MW. Construct validity of the WISC-III verbal and performance factors for Black special education students. *Assessment* 1997;4:297–304.
- Loewenstein DA, Arguelles T, Arguelles S, Linn-Fuentes P. Potential cultural bias in the neuropsychological assessment of the older adult. *Journal of Clinical and Experimental Neuropsychology* 1994;16(4):623–629. [PubMed: 7962363]
- Manly JJ, Jacobs DM, Sano M, Bell K, Merchant CA, Small SA, Stern Y. Cognitive test performance among nondemented elderly African Americans and whites. *Neurology* 1998;50(5):1238–1245. [PubMed: 9595969]
- Manly JJ, Jacobs DM, Touradji P, Small SA, Stern Y. Reading level attenuates differences in neuropsychological test performance between African American and European American elders. *Journal of the International Neuropsychological Society* 2002;8(3):341–348. [PubMed: 11939693]
- Marcopulos BA, McLain CA, Giuliano AJ. Cognitive impairment or inadequate norms: A study of healthy, rural, older adults with limited education. *The Clinical Neuropsychologist* 1997;11:111–131.
- Marsh HW, Balla JR, McDonald RP. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin* 1988;103(3):391–410.
- Marsiske M, Klumb P, Baltes MM. Everyday activity patterns and sensory functioning in old age. *Psychology and Aging* 1997;12:444–457. [PubMed: 9308092]
- Mast BT, Lichtenberg PA. Assessment of functional abilities among geriatric patients: A MIMIC model of the functional independence measure. *Rehabilitation Psychology* 2000;45(1):49–64.
- Mast BT, MacNeill SE, Lichtenberg PA. A MIMIC model approach to research in geriatric neuropsychology: The case of vascular dementia. *Aging, Neuropsychology, and Cognition* 2002;9(1):21–37.
- Miles, GT. Neuropsychological assessment of African Americans. In: Ferraro, FR., editor. *Minority and cross-cultural aspects of neuropsychological assessment*. Lisse, Netherlands: Swets & Zeitlinger Publishers; 2002. p. 63–77.
- Muthen BO. Latent variable modeling in heterogeneous populations. *Psychometrika* 1989;54:557–585.
- National Center for Health Statistics. *Current Estimates from the National Health Interview Survey, 1989*. Hyattsville, MD: Public Health Service; 1990.
- Patton DE, Duff K, Schoenberg MR. Performance of Cognitively Normal African Americans on the RBANS in Community Dwelling Older Adults. *The Clinical Neuropsychologist* 2003;17(4):515–530. [PubMed: 15168916]
- Rey A. L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie* 1941;28:21.
- Ripich DN, Carpenter B, Ziolo E. Comparison of African-American and white persons with Alzheimer's disease on language measures. *Neurology* 1997;48:781–783. [PubMed: 9065567]
- Rosano C, Simonsick EM, Harris TB, Kritchevsky SB, Brach J, Visser M, Yaffe K, Newman AB. Association between physical and cognitive function in healthy elderly: the health, aging and body composition study. *Neuroepidemiology* 2005;24(1–2):8–14. [PubMed: 15459503]
- Saenz TI, Huer MB. Testing strategies involving least biased language assessment of bilingual children. *Communication Disorders Quarterly* 2003;24(4):184–193.
- Schaie, KW. Schaie – Thurstone adult mental abilities test manual. Palo Alto, CA: Consulting Psychologists Press; 1985.

- Skiba RJ, Knesting K, Bush LD. Culturally competent assessment: More than nonbiased tests. *Journal of Child and Family Studies* 2002;11(1):61–78.
- Teresi JA, Holmes D, Ramírez M. Performance of cognitive tests among different racial/ethnic and education groups: Findings of differential item functioning and possible item bias. *Journal of Mental Health and Aging* 2001;7(1):79–89.
- Whitfield KE. Studying cognition in older African Americans: Some conceptual considerations. *Journal of Aging and Ethnicity* 1996;1(1):35–45.
- Whitfield KE. Challenges in cognitive assessment of African Americans in research on Alzheimer's disease. *Alzheimer's Disease & Associated Disorders* 2002;16 (S2):S80–S81.
- Whitfield KE, Baker-Thomas T, Heyward K. Evaluating a measure of everyday problem solving for use in African Americans. *Experimental Aging Research* 1999;25(3):209–221. [PubMed: 10467512]
- Whitfield KE, Fillenbaum GG, Pieper C, Albert MS, Berkman LF, Blazer DG, Rowe JW, Seeman T. The effect of race and health-related factors on naming and memory: The MacArthur studies of successful aging. *Journal of Aging & Health* 2000;12(1):69–89. [PubMed: 10848126]
- Whitfield KE, Weidner G, Clark R, Anderson NB. Sociodemographic diversity and behavioral medicine. *Journal of Consulting and Clinical Psychology* 2002;70(3):463–481. [PubMed: 12090363]
- Whitfield KE, Wiggins S. The impact of desegregation on cognition among older African Americans. *Journal of African American Psychology* 2003;29(3):275–291.
- Whitfield KE, Willis S. Conceptual Issues and analytic strategies for studying cognition in older African Americans. *African-American Research Perspectives* 1998;4(1):115–125.
- Wilson, BA.; Cockburn, J.; Baddeley, AD. *The Rivermead Behavioral Memory Test*. Bury St. Edmunds, UK: Thames Valley Test Co; 1985.
- Unverzagt FW, Hall KS, Torke AM, Rediger JD. Effects of age, education, and gender on CERAD neuropsychological test performance in an African American sample. *The Clinical Neuropsychologist* 1996;10:180–190.
- U.S. Census Bureau. Table 3. Educational Attainment of the Population 15 Years and Over, by Marital Status, Age, Sex, Race, and Hispanic Origin. 2000.
<<http://www.census.gov/population/socdemo/education/ppl-169/tab03.xls>> published March 2002
- Vincent KR. Black/White IQ differences: Does age make a difference? *Journal of Clinical Psychology* 1991;47:266–270. [PubMed: 2030133]
- Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* 1992;30:473–483. [PubMed: 1593914]
- Zsembik BA, Peek MK. Race differences in cognitive functioning among older adults. *Journal of Gerontology: Social Sciences* 2001;56(B-5):S266–S274.

Figure 1. MIMIC Model: An illustration

Note. *a* paths are factor loadings, and express the relationship between the latent construct and each of three measures of that construct; the *b* path shows the relationship between the exogenous covariate and the latent construct. This path addresses whether there is an overall relationship between the covariate and the construct. The *c* path assesses whether, after controlling for the relationship between covariate and the construct, there is an *additional* unique direct residual relationship between the covariate and each indicator. If the relationship between the covariate and Measure 1 is *more strongly positive* than the relationship between the covariate and the construct, then one would expect an additional significant direct positive path between the covariate and Measure 1. This would suggest that Measure 1 is positively biased. In other words, it shows a significantly stronger relationship to the covariate than the overall construct. Expressed differently, this would mean that the residual variance in Measure 1 (variance not explained by the common construct factor) shows additional relationship to the covariate, above and beyond that covariate's indirect influence on Measure 1 via the latent construct.

Figure 2. MIMIC Model in the current study

Note. MIMIC model of race effects on reasoning, memory, and speed, with years of education, age and gender as covariates. Paths shown are as follows: *a*, influence of ethnicity on latent variable; *b*, influence of covariates on latent variable; *c*, influence of race on specific measure; *d*, influence of covariates on specific measure; *e*, factor loading (relationship between latent construct and specific measure). **Paths are shown only for one latent factor, but these paths were estimated for all latent variables and measured indicators.** Dummy variables representing each of the assessment sites were also included as exogenous covariates in these models, but they are not shown in order to simplify the figure.

Table 1

Characteristics of the ACTIVE sample at baseline.

	Total Sample (N = 2741)		African Americans (n = 729)		Whites (n = 2012)		<i>Cohen's d</i>	<i>t</i> ² _k	df	<i>p</i>
	Mean	SD	Mean	SD	Mean	SD				
Age	73.6	5.9	72.1	5.3	74.1	6.0	-0.35	8.230	1441	<.001
Years of Education	13.5	2.7	13.0	2.7	13.7	2.7	-0.26	6.574	1292	<.001
Gender (% female)	75.8	-	83.3	-	73.2	-	-	29.82	1	<.001
Physical Function	68.8	24.1	65.0	24.6	70.2	23.7	-0.22	4.952	1224	<.001

Note: T-tests show significance of the comparison between African American and White participants; for Gender, a corresponding chi-squared test was used. No covariate adjustment was employed.
¹ Degrees of freedom and t-statistic were adjusted for non-homogeneity of variance.

Table 2**Cognitive Measures**

Domain	Measure	Description
Reasoning	Word Series Test; Schaie – Thurstone Adult Mental Abilities Test (Schaie, 1985).	Assessed the ability to correctly determine how a series of words(e.g., days of the week, months of the year) progress. There were 30 multiple choice items
	Letter Series Test; Schaie – Thurstone Adult Mental Abilities Test (Schaie, 1985).	Assessed the ability to correctly identify the letter that appropriately completes a series of letters for each of the test items. There were 30 multiple-choice items.
	Letter Sets Test; Kit of Factor Referenced Cognitive Tests (Ekstrom et al., 1976).	Assessed the ability to identify the rule that related four groups of letters to each other, and subsequently identify a fifth group of letters to which the rule did not apply. There were 15 items.
Memory	Rey Auditory Verbal Learning Test; (Rey, 1964)	Assessed the ability to form and retain new verbal memories for a list of 15 semantically unrelated words. There were five trials of learning and recall.
	Hopkins Verbal Learning Test – Revised; (Benedict Schretlen, Groninger, & Brandt, 1998)	Assessed the ability to form and retain new verbal memories for a list of 12 semantically related words. There were three trials of learning and recall. A recognition trial was also given, but is not included in this study.
	Rivermead Behavioural Memory Test; (Wilson, Cockburn, & Baddeley, 1985).	Assessed the ability to learn and retain verbal memories for a brief story of 21 propositions. A free recall trial was given.
Processing Speed	Useful Field of View Test (UFOV); (Ball & Owsley, 1993)	Assessed elements of an individual's field of visual attention: simple perceptual identification (UFOV1: Speed), perceptual identification and localization of peripheral target (UFOV2: Divided Attention), identification and localization of peripheral object in visual clutter (UFOV3: Selective Attention), and perceptual comparison and peripheral localization in clutter (UFOV4: Selective Attention). Scores on each subtask represented the minimum presentation time needed for 75% accuracy in

Table 3

Mean cognitive test performance of the ACTIVE sample at baseline (controlling for demographic variables and data collection site).

	Total Sample (N = 2741)		African Americans (n = 729)		Whites (n = 2012)		Cohen's d	t	df	p
	Mean	SD	Mean	SD	Mean	SD				
Word Series	9.49	4.89	7.21	4.19	10.32	4.86	-0.69	16.39	1483.74 [/]	<.001
Letter Series	10.04	5.55	7.17	4.82	11.08	5.43	-0.76	18.12	1440.49 [/]	<.001
Letter Sets	5.75	2.79	4.87	2.65	6.07	2.77	-0.44	10.15	2739.00	<.001
AVLT	48.51	10.59	45.92	10.53	49.45	10.45	-0.34	7.81	2739.00	<.001
HVLT	26.11	5.49	24.64	5.57	26.64	5.36	-0.37	8.56	2739.00	<.001
Rivermead	6.29	2.75	5.39	2.40	6.62	2.80	-0.47	11.31	1490.59 [/]	<.001
UFOV 1	30.81	40.76	34.01	45.95	29.65	38.66	0.10	-2.29	1122.62 [/]	.022
UFOV 2	133.28	124.86	156.91	134.50	124.72	120.07	0.25	-5.69	1173.66 [/]	<.001
UFOV 3	320.64	134.16	341.54	133.89	313.07	133.49	0.21	-4.93	2739.00	<.001
UFOV 4	456.75	68.55	464.80	66.58	453.83	69.03	0.16	-3.71	2739.00	<.001

Note: AVLT = Rey Auditory Verbal Learning Test, HVLT = Hopkins Verbal Learning Test, Rivermead = Rivermead Paragraph Recall Task, UFOV1-UFOV4 = Useful Field of View subtasks 1 through 4.

[/] Degrees of freedom and t-statistic were adjusted for non-homogeneity of variance.

Invariance tests for proposed three-factor model of cognition (N = 2,741; *n* = 729, African Americans; *n* =2012, Whites).

Table 4

	χ^2 (df)	<i>p</i>	χ^2 Difference	<i>p</i>	NFI	IFI	RMSEA	CFI	Model AIC
Model One	444.17 (77)	< .001	-	-	0.96	0.97	0.042	0.97	550.17
Model Two	421.34 (71)	< .001	22.83 (6)	.0009	0.97	0.97	0.042	0.97	539.34
Model Three	410.67 (64)	< .001	10.67 (7)	.154	0.97	0.97	0.044	0.97	542.67

Note: In Model One, both unstandardized factor loadings and factor covariances were constrained to equality across groups. In Model Two, only unstandardized factor loadings were constrained to equality across groups. In Model Three, configural invariance (i.e., same pattern of factor loadings in both groups) was tested. NFI= Normed Fit Index; IFI= Incremental Fit Index; RMSEA= Root Mean Square Error of Approximation; CFI= Comparative Fit Index.

Table 5

Completely standardized factor loadings of cognitive measures on hypothesized latent factors at baseline assessment occasion (N = 2,741; *n* = 729, African Americans; *n* = 2012, Whites).

	Reasoning	Memory	Processing Speed
Word Series	.91**		
Letter Series	.91**		
Letter Sets	.69**		
AVLT		.79**	
HVLT		.83**	
Rivermead		.59**	
UFOV 1			.53**
UFOV 2			.73**
UFOV 3			.83**
UFOV 4			.61**

Note: AVLT = Rey Auditory Verbal Learning Test, HVLT = Hopkins Verbal Learning Test, Rivermead = Rivermead Paragraph Recall Task, UFOV1-UFOV4 = Useful Field of View subtasks 1 through 4;

*
p < .05,

**
p < .01,

p < .001

Table 6

MIMIC Model Results: Standardized regression coefficients between exogenous variables and latent factors (N = 2,741)

	Reasoning	Memory	Speed
African American	-.36***	-.26***	.15***
Covariates			
Education	.38***	.28***	-.16***
Age	-.36***	-.39***	.50***
Gender	.05**	.29***	.00
Physical Function	.08***	.12***	-.10***

Note:

*
p < .05,

**
p < .01,

p < .001

MIMIC Model Results: Standardized regression coefficients between exogenous variables and measured indicators, above and beyond latent variable effects ($N = 2,741$).

Table 7

	African American	Education	Age	Gender	Physical Function
Reasoning					
Word Series	0.01	-0.01	-0.01	0.01	0.01
Letter Series	-0.01	0.00	0.01	0.00	-0.03
Letter Sets	0.03	0.03	0.00	-0.03	0.03
Memory					
AVLT	0.03	-0.02	-0.02	0.05**	0.01
HVLT	-0.02	0.01	0.02	0.00	0.00
Rivermead	-0.05**	0.06**	-0.01	-0.17***	-0.03
Processing Speed					
UFOV 1	0.01	-0.01	0.06**	-0.02	-0.03
UFOV 2	0.05*	-0.02	0.02	0.02	-0.01
UFOV 3	-0.01	0.01	-0.02	-0.01	0.01
UFOV 4	-0.04	0.03	-0.01	0.01	-0.01

Note:
* $p < .05$,
** $p < .01$,
*** $p < .001$