Considering Measurement Bias in Cognition at the Intersection of Race and Gender

by

Julia Keen Goodwin

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Sociology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: December 6, 2023

The dissertation is approved by the following members of the Final Oral Committee:

Katherine J. Curtis (co-chair), Professor, Community & Environmental Sociology
Mosi Adesina Ifatunji (co-chair), Assistant Professor, Department of African American Studies and Sociology
Jason Fletcher, Professor, LaFollette School of Public Affairs
Monica Grant, Professor, Department of Sociology
Lauren Schmitz, Assistant Professor, LaFollette School of Public Affairs

ABSTRACT

Race and gender disparities in the prevalence and incidence Alzheimer's Disease and related dementias (ADRDs) are well-documented in national reports: In the U.S., Black people are more likely to have ADRDs than White people, while women make up the majority of those aged 65 and older with ADRDs. Scholars in the social sciences frequently utilize large, nationally-representative, longitudinal surveys that collect detailed data on cognition and respondents' lives to identify the influence that modifiable risk factors have on race and gender group differences. However, the ability of these commonly explored individual and contextual variables to fully explain cognition disparities by race and gender are limited. Education, socioeconomic status (SES), stress, and geography fail to fully account for the disparities between Black and White samples and between men and women. A central question endures: What accounts for the remaining race- and gender-based disparities in cognition that social contextual factors cannot explain?

The majority of research detailing race and gender disparities in cognition are conducted under the assumption—without demonstration—of *measurement invariance*, or the essential characteristic that latent variables quantify the same underlying concept across cultures, societies, geographies, and time. Like other assessments that attempt to capture a universal measure of intelligence or aptitude (e.g., IQ test, SATs), cognition assessments in these large, longitudinal surveys may have questions that are unintentionally biased towards certain groups. For example, especially for aging adults, an item that asks people to count backwards from 100 by 7s may be easier for men compared to women because of gendered expectations and institutional steering: men have historically been stereotyped as being "good at math" and encouraged to build that skill, while women have not been encouraged in the same way. The item

would be considered biased towards men if, on average, despite both groups having similar cognition, men answered the question correctly and women answered incorrectly. The purpose of the following dissertation is to shift the focus towards the cognition assessment itself, asking to what extent does bias in the measurement of cognition contribute to the persistent disparities in cognition across intersectional race-gender groups.

The following dissertation fits into larger historical and current efforts to accurately measure cognition across diverse populations within the United States and cross-nationally. Of particular interest in this study is the measurement of cognition in the Health and Retirement Study (HRS): previous research has established that the cognition assessment in the HRS exhibits measurement invariance separately across race—between non-Hispanic Black, non-Hispanic White, and Hispanic—and gender—between men and women. However, researchers have not examined to what extent measurement invariance holds at the intersection of race and gender in the HRS. In line with the HRS's efforts to improve the accuracy of the cognition assessment, the following analyses further contemplates measurement invariance by intersectional identity, attending to potentially neglected disparities in the measurement of cognition. In Chapter 2, I perform a multigroup confirmatory factor analysis (MGCFA) of cognition in a cross-sectional sample of the Health and Retirement Study (HRS) across four intersectional race-gender groups, which estimates the presence and degree of bias in the cognition measures. I find evidence for measurement non-invariance, specifically that the underlying cognition factor is not captured equivalently for Black women and White women, as compared to Black men and White men. These findings demonstrate the importance of ensuring measurement invariance for latent variables across intersectional groups.

In the broader social sciences and specific field of demography, scholars have endeavored to understand and explain differences in cognition trajectories—and not just in cross-sectional samples—by race and gender in old age. In Chapter 3, using data from 1996 to 2018 in the HRS, I calculate a second-order latent growth model, which permits me to calculate the amount of measurement invariance in cognition over age. In addition, I borrow from demographic research on lifespan variation to conceptualize and hypothesize how variation in cognition trajectories behave across intersectional race-gender groups. Adjusting the measurement model in the latent growth model in the cognition age-trajectories of Black women, Black men, White women, and White men results in a 50% reduction in the gap between Black women and White men. Additionally, I found that there was greater variation in the trajectories of Black women and Black men even when measurement bias was removed, suggesting that White women and men experience greater certainty in their cognition trajectories in old age. This novel modeling strategy demonstrates the importance of assessing and correcting for bias in the measurement of cognition trajectories across race-gender groups and considering the often-overlooked influence of variation in cognition trajectories to race-gender disparities in cognition.

Lastly, in Chapter 4, I add to the extant research by removing measurement bias from cognition age-trajectories while simultaneously controlling for three relevant covariates. Using data from Health and Retirement Study (HRS), I calculate an unbiased second-order latent growth model while also controlling for allostatic load to measure individual levels of chronic stress; years of education; and the average years lived in the Southern United States. Education had the largest influence on the level, but not change, of cognition across all four intersectional race-gender groups. When I control for all three variables, the disparities in cognition trajectories between Black women, Black men, and White men are substantively eliminated, while White

women's cognition remains at a comparatively high level. These analyses illustrate the potential explanatory power of measurement invariance analyses for explaining the remaining gaps in cognition trajectories.

Taken together, these three empirical studies advance scholarship on measurement invariance in cognition assessments by utilizing an intersectional lens to reveal sizeable bias across race-gender groups that would have gone otherwise overlooked. These analyses consider measurement invariance within an intersectional framework, which is important to examine if we, as a research community, intend to understand the reliability and validity of cognition measurement and improve the lives of aging people in the United States and globally. These results suggest that scholars who study disparities in cognition—or any complex, latent facet of health and well-being—must consider how bias may be contributing to disparities.

ACKNOWLEDGEMENTS

I owe a debt of gratitude to the people who helped me complete this dissertation and navigate graduate school. Katherine Curtis has been an incredible co-advisor throughout graduate school, including helping me with my master's paper and this dissertation, starting in its infancy to its completion. She has been generous with her time, offering helpful and thoughtful feedback, and always acknowledging the importance of being present. Mosi Ifatunji has been tremendously helpful as my co-advisor and has also been generous with his time, helping move this project forward. What makes working with Mosi especially delightful is that his energy is infectious. Every meeting with him left me invigorated to go out and do research. I also want to thank Monica Grant for her flexibility, responsiveness, and thoughtful feedback before and during my dissertation defense. Lauren Schmitz and Jason Fletcher provided excellent perspectives and feedback that truly improved my dissertation. Doug Hemeken and Jason Struck of the SSCC were incredibly helpful in assisting me in learning a new statistical program and methodology for this dissertation. I appreciate all their help and time, and their capabilities to take on the challenge of working in Mplus. I also want to acknowledge the staff in the Sociology department and the CDE who assisted me in navigating the administrative parts of graduate school.

I had the great privilege of being surrounded by thoughtful, smart colleagues and friends while at the University of Wisconsin, so I want to thank all the CDE students who have made Dem Sem and the CDHA seminars a pleasure to attend and who made the offices on the 8th and 4th floor a great place to work. I want to especially thank my colleagues Kelsey Wright and Elisa Avila for their perspectives, support, and friendship.

I also want to thank my family for always being so supportive throughout this endeavor. There's nothing quite like housing your 35-year-old daughter while she writes her dissertation, so I want to especially thank my mom, Sue, and stepdad, Lou, for being patient, listening, and being supportive throughout graduate school. I'd like to separately thank my mom for teaching me to think critically, to ask questions, and to not be afraid to be a squeaky wheel. I owe a great deal of my sanity to her and her patience. I want to thank Lou for his unending support and kind words. Despite being (I would say, an appropriately) moody and obstinate teen when we first met, Louis has been an incredible father to me. To my oldest sister Katie for being my surrogate mother early on (I don't think I would be here if you hadn't potty trained me) and later for being a friend. To my next eldest sister Alice, I want to thank you for not trying to kill me *The Good Son*-style and giving us time to become friends. I want to thank both of my sisters for shaping me into the woman I am today. I also want to thank Andy, little Alice, and Hannah for making being an aunt a joy.

I truly would not have made it without my friends for their encouragement and support. To Luke, for being a partner in trying new restaurants, exploring new cities, and appreciating the arts. I also want to thank my DC-area friends—Charlotte, Aditi, Lily, Becky, Chloe, and Maddie—for their unquestioning support, I don't think would have survived dissertation writing without it.

Table of Contents

CHAPTER 1. INTRODUCTION

Race and gender disparities in the prevalence and incidence Alzheimer's Disease and related dementias (ADRDs) are well-documented in national reports: In the United States, Black people fair worse than their White counterparts in the prevalence and incidence of Alzheimer's disease and related dementias (ADRDs), while women make up two-thirds of those over the age of 65 with ADRDs (Rajan et al. 2019). These enumerations reflect a pressing public health concern. As the population continues to age (Vespa, Medina, and Armstrong 2020) and the composition of the aging population continues to change (U.S. Census Bureau 2018), the burden of ADRDs will fall squarely on the backs of more Black people than White people and more women than men (Rajan et al. 2021). Experts have made a concerted effort to explain race and gender differences in ADRDs by identifying modifiable risk factors. These are the social, contextual, and behavioral factors that can be controlled or changed and could potentially prevent, delay, and slow the debilitating effects of dementia while also reducing disparities in ADRDs (Livingston et al. 2020; Nianogo et al. 2022).

Social scientists frequently utilize large, nationally-representative, longitudinal surveys that collect data on cognition to explain these race- and gender-based disparities by controlling for social and contextual factors that systematically differ by race and gender. Because race- and gender- based cognition disparities are not, by definition, essential to socially constructed social statuses, the principal drivers of disparities are the larger structural factors that produce these disparities (Braveman et al. 2011; Braveman and Gottlieb 2014; Williams 2012; Williams and Mohammed 2013; Williams and Sternthal 2010). Structural forces differentially distribute access to resources that improve cognition in old age, or conversely, distribute exposures to harm that hinder cognition. Education, employment, state of residence, and incidences of racism and

sexism are but a few important social factors that contribute the race and gender disparities in cognition. Though controlling for these variables reduces disparities, wide race and gender gaps in cognition remain (Byrne and Anaraky 2022; Chen et al. 2022; Cintron et al. 2023; Hayward et al. 2021). A central question endures: What accounts for the remaining race- and gender-based disparities in cognition that social contextual factors cannot explain?

An answer lies in assessing whether or not and to what extent cognition assessments exhibit measurement invariance. Measurement invariance is a foundational assertion that the latent variable of interest—here, cognition—is captured by the same set of tests and questionnaire items over social dimensions and time (Horn and McArdle 1992). Much like purportedly universal assessments of intelligence (e.g., IQ tests) or scholastic aptitude (e.g., SATs, ACTs, GREs), there may be items in cognition assessments that do not reflect actual cognitive ability, but inadvertently only capture, for example, level of education. Cognition assessments in large, longitudinal surveys may have questions that are unintentionally biased towards certain groups as a result of institutional discrimination and/or societal norms. For example, especially in samples of older Americans, an item that asks people to count backwards from 100 by 7s may be easier for men compared to women as a result of gendered expectations and institutional steering: men have historically been stereotyped as being "good at math" and encouraged to build that skill, while women have not been encouraged in the same way. This item would be considered biased towards men if men systematically answered the question correctly while women systematically answered incorrectly, despite both groups having similar levels of cognition.

When a survey item does not capture the same underlying construct across socially constructed race and gender statuses, the item is considered non-invariant, or biased (Horn and

McArdle 1992). Thus, some proportion of the differences in cognition may not be real differences in cognitive abilities but are an artifact of measurement non-invariance or bias. This dissertation fits into larger efforts—historically and currently—to accurately measure cognition and diagnose dementia across diverse populations within the United States and internationally. In recent decades, the United States has invested heavily in enumerating not only the prevalence of dementia in aging Americans, but the number of people with pre-clinical mild cognitive impairment without dementia (cognitive impairment, no dementia or CIND) (Langa et al. 2005, 2020). These censuses, carried out through federally-funded longitudinal survey research, are important given the rapidly aging population and the desire to diminish the debilitating effects of Alzheimer's disease. Importantly, populations in low- and middle-income countries of the Global South are also rapidly aging, making it imperative to understanding the etiology of dementia and CIND outside Western contexts to be able to make cross-national comparisons (Gross et al. 2023; Langa et al. 2020).

An important contribution of the following analyses is utilizing intersectionality as a framework to hypothesize about the presence and extent of measurement invariance in cognition assessments. A central tenant of intersectionality is that social statuses—race, ethnicity, gender, class, sexuality, nationality—do not exist separately, but are mutually constitutive (Collins 2015; Crenshaw 1989). The original impetus of intersectionality was to highlight how the U.S. legal system erases Black women's experiences of work discrimination and violence by refusing to simultaneously consider race and gender (Crenshaw 1989, 1991). Social scientists have utilized intersectionality to reframe their understanding of power structures and inequality to consider how existing at the nexus of multiple, simultaneously existing identities influences outcomes in health and well-being (Bauer 2014; Bauer and Scheim 2019). I focus on race and gender,

specifically disaggregating my analytical samples into four categories—Black women, Black men, White women, and White men—to better understand how cognition is assessed among people with these intersecting statuses in the United States. The main thrust of these analyses is to add to the scholarship endeavoring to understand measurement in diverse contexts and to be able to make comparisons across numerous, intersecting social categories.

The purpose of the following dissertation is to shift the focus towards the cognition assessment itself, asking to what extent does bias in the measurement of cognition contribute to the well-established disparities in cognition across race and gender. Across the three chapters, I use the Health and Retirement Study (HRS) from 1996 to 2018 using the survey's TICS assessment as the primary outcome variable. In Chapter 2, I quantify the amount of bias in the latent variable of cognition in a cross-sectional sample of adults aged 50 years and older in the HRS for Black women, Black men, White women, and White men. Chapter 3 asks if and to what extent there is bias in the measurement of cognition longitudinally over the four intersectional race-gender groups by utilizing a latent growth model. In Chapter 4, I contribute to the extant literature by removing bias in cognition due to measurement non-invariance and control for three social, contextual factors that social scientists have found to influence trajectories of cognition in old age: allostatic load (AL), education, and Southern residence. The concluding chapter outlines the findings of the dissertation and potential directions for future research. The central takeaway of my dissertation is that researchers must consider the impact of measurement bias when comparing cognition across socially-defined groups. Importantly, this consideration must go beyond the impact of measurement bias on a single social status, but at the intersection of multiple social statuses. Indeed, previous research that assesses measurement invariance separately by race and gender come to the conclusion that cognition in the HRS is unbiased

across race groups and gender groups (e.g., Blankson and McArdle 2015). However, as I demonstrate in my dissertation, there is bias in the HRS's cognition assessment, but at the intersection of race and gender in cross-sectional and longitudinal analyses.

CHAPTER 2. MEASUREMENT INVARIANCE OF COGNITION AT THE INTERSECTION OF RACE AND GENDER IN THE HRS: REMOVING BIAS INCREASES BLACK AND WHITE WOMEN'S AVERAGE COGNITION SCORES RELATIVE TO MEN

ABSTRACT

Previous research has established that cognition measurements in the Health and Retirement Study (HRS) and similar longitudinal studies of older Americans exhibit *measurement invariance*, or the essential characteristic that latent (or indirectly measured) variables quantify the same underlying concept separately across race (between non-Hispanic Black, non-Hispanic White, and Hispanic samples) and gender (between men and women) (Blankson and McArdle 2015; McArdle 2011; McArdle, Fisher, and Kadlec 2007). However, researchers have not examined to what extent measurement invariance holds at the intersection of race and gender. By neglecting measurement invariance by intersectional identity, researchers may overstate disparities in cognition and risk misspecifying or misinterpreting results. In this chapter, I perform a multigroup confirmatory factor analysis (MGCFA) of cognition in the HRS across intersectional race-gender groups, which estimates the presence and degree of bias in the cognition measures. I find evidence for measurement non-invariance, specifically that the underlying cognition factor is not captured the same way for Black women and White women, as compared to Black men and White men. Black women had a substantial change in their average cognition, with a 0.483 unit increase in their average cognition after I adjusted the measurement model. White women's mean value also saw a significant change, increasing by 0.439 units. Black men had a modest increase in their cognition estimate, increasing by 0.144 units. These findings demonstrate the importance of ensuring measurement invariance for latent variables across intersectional groups.

INTRODUCTION

In the United States, race and gender shape cognition in old age, though less often researchers estimate cognition at the intersection of these two social dimensions. By any measure, non-Hispanic Whites (referred to as White people hereafter) have more favorable outcomes for Alzheimer's and related dementias (ADRDs) and cognition compared to non-Hispanic Blacks (Black people, hereafter) (Alzheimer's Association 2022). Census estimates (Rajan et al. 2021; U.S. Census Bureau 2021), incidence rates (Steenland et al. 2016; Weuve et al. 2018), and estimated lifetime risk (Power et al. 2021) of ADRDs all point to a race-cognition gradient that advantages the White over the Black population. Differences by gender are slightly more complicated: women make up nearly two-thirds of ADRD cases in the United States (Rajan et al. 2021), though this may be attributed to women's longevity, and thus greater lifetime risk, and not due to sex per se (Chêne et al. 2015; Fitzpatrick et al. 2004; Hale, Schneider, Gampe, et al. 2020; Hebert et al. 2001; Levine et al. 2021). Experts have made a concerted effort to explain race and gender differences in ADRDs by identifying modifiable risk factors—like education, discrimination, income, employment—by analyzing large, nationally-representative, longitudinal surveys to control for these variables. Though controlling for these risk factors reduces disparities, wide race and gender gaps in cognition still remain (Byrne and Anaraky 2022; Chen et al. 2022; Cintron et al. 2023; Hayward et al. 2021). These data beg the question: what accounts for the race and gender differences in cognition old age?

Most research on race and gender differences in cognition is executed under the assumption of measurement invariance. Measurement invariance is a foundational assertion that the latent variable of interest—here, cognition—is captured by the same set of tests and questionnaire items over social dimensions and time (Horn and McArdle 1992). However, it may

be that some proportion of the differences in cognition are not real differences in cognitive abilities but an artifact of measurement non-invariance, where survey items and tests do not capture the same underlying construct across sub-groups (Horn and McArdle 1992). Take, for example, any popular standardized test, like the SATs or IQ tests: these tests are attempting to assess some universal measure of cognitive ability or aptitude, but previous analyses find that race and social class are the best predictors of success on these tests (Wicherts 2016; Wicherts and Dolan 2010; Zwick 2019). The same goes for cognition assessments: it would be problematic to suggest, after looking at the disparities by race and gender, that these differences are due to some inherent deficiency. Instead, the questions may not capture a universal measure of cognition, but the ways that societal expectations and norms shape people's cognitive abilities. For example, especially for aging adults, an item on a cognition assessment in a survey that asks people to count backwards from 100 by 7s may be easier for men compared to women because men have historically been stereotyped as being "good at math" and encouraged to pursue STEM field in school, while Women may not have been given the same encouragement, and thus have not had the same expectation around math. This item would be considered biased towards men if men answered the question correctly and women answered incorrectly, despite both groups having similar levels of cognition. The item would thus be indicative of historical marginalization in education, and not cognition per se. Thus, some proportion of the group differences are due to mismeasurement in the latent variable itself.

Bias due to measurement non-invariance in cognition assessments may lead researchers to incorrectly evaluate race and gender disparities in cognition. In addition, not only is there potential for biased results by race and gender categories separately, but at the intersection of those identities. It is imperative to consider intersectional groups when estimating potential bias

so that researchers can better understand true cognitive abilities among those most at risk for ADRDs and develop more universal assessments of cognition. Measurement invariance analyses offers a solution to identifying and correcting for biases in cognition assessments and any erroneous conclusions that may follow. In this study, I perform measurement invariance analyses in order to assess whether cognition items in the Health and Retirement Study (HRS) reflect the same underlying construct across race-gender intersectional groups. I will determine whether the items in the HRS's cognition series contribute equally to the measurement of the latent variable—cognition—by intersectional sub-group. If measurement invariance does not hold, I will identify the sources of non-invariance (i.e., specific questionnaire items), and estimate the proportion of the intersectional group differences that are due to non-invariance, and calculate new intersectional sub-group means. These new summary statistics will reflect more accurate levels of cognition for each intersectional sub-group by removing the proportion of that difference that is due to measurement non-invariance. Without ensuring measurement invariance, researchers fail to take into account the reality that cognition assessments—like any assessment—has the potential to be biased towards marginalized social groups. Unadjusted results have the potential to further add to the White supremacist notion that White people have higher cognition that Black people, when in reality the differences are far more complex, especially when viewed through an intersectional lens.

BACKGROUND

Cognition Disparities in Aging Populations

Cognition is a growing concern in the United States given its rapidly aging and changing population. Projections from the U.S. Census estimate that by 2030, one out of five Americans

will be over the age of 65 (Vespa et al. 2020) and by 2034, for the first time ever, there will be more adults over the age of 65 than children under 18 years old (Vespa et al. 2020). In addition, the race and gender composition of people over 65 is projected to change significantly: census projections predict that over the next 40 years, the elderly population will have a larger proportion of Black and Latine people and a smaller proportion of White people, while women are expected to continue to outlive men, though the sex ratio gap in older ages is projected to decrease over time (U.S. Census Bureau 2018; Vespa et al. 2020). These demographic changes in the aging population are important to understand given the wide disparities in cognitive decline and ADRDs. As a result of the rapidly aging population, the prevalence of ADRDs is projected to grow by 60% between 2020 and 2050, barring any major medical breakthroughs to prevent, slow, or cure ADRDs (Alzheimer's Association 2020).

Given current disparities in ADRDs by race and gender—save for major public health interventions—the burden of ADRDs will fall disproportionately on Black people relative to White people, and more women than men. According to estimates using the 2020 Census, 18.6% of Black people are estimated to have Alzheimer's disease compared to 10.0% of White people (Rajan et al. 2021; U.S. Census Bureau 2021). Incidence rates of ADRDs over the last ten years are between 1.6 to 2.0 times higher for Black people compared to White people (Steenland et al. 2016; Weuve et al. 2018), while lifetime risk for ADRDs among Black people is 40-80% higher compared to White people (Power et al. 2021). Examining differences by gender, women make up nearly two-thirds of Alzheimer's disease cases in the United States, while approximately 60% of Alzheimer's patients over 65 years old are women (Rajan et al. 2021). Though women have an increased lifetime risk of developing an ADRD compared to men, researchers attribute these gender disparities to women's longevity and not due to sex per se (Chêne et al. 2015; Hebert et

al. 2001). Other findings support this assertion: age-adjusted incidence and prevalence rates are similar between men and women (Fitzpatrick et al. 2004; Hale, Schneider, Gampe, et al. 2020; Levine et al. 2021; Rajan et al. 2021), suggesting women's longer life expectancy increases the risk of ADRDs.

The wide racial and gender disparities in ADRDs in conjunction with the rapidly aging population and its changing demographics suggests that we must better understand the social contexts that contribute to differential cognitive decline to ensure that future aging populations—and their caregivers—are provided the financial resources and medical care to ensure their well-being. Disparities sourced from national, official enumerations reflect a pressing public health concern. Because race- and gender- based cognition disparities are not, by definition, essential to socially constructed social statuses, the principal drivers of disparities are the larger structural factors that produce these disparities (Braveman et al. 2011; Braveman and Gottlieb 2014; Williams 2012; Williams and Mohammed 2013; Williams and Sternthal 2010). Social scientists have made a concerted effort to explain race and gender differences in ADRDs by identifying modifiable risk factors, or the social, contextual, and behavioral factors that can be controlled or changed and potentially lead to ADRD prevention and reduction (Livingston et al. 2020; Nianogo et al. 2022). Scholars in the social sciences frequently utilize large, nationally-representative, longitudinal surveys that collect data on cognition to estimate group differences and identify the influence that modifiable risk factors have on cognition. Cognition assessments in large surveys, like the Health and Retirement Study (HRS), are not necessarily meant to diagnose Alzheimer's disease or dementia, but to surveille levels and changes in cognition over age and across socially defined groups.

Cognition is an important facet of health in the later years of life. Though subtle changes in cognition in old age are expected as early as age 60 (Salthouse 2019), these small declines usually do not bar people from being able to continue to take care of themselves (perhaps with assistance), enjoy leisure activities, and have meaningful social relationships. The smaller, more subtle changes in an individual's cognition, like memory, may not be clinically actionable, but are important to investigate in population surveys to better understand how modifiable risk factors contribute to levels and changes in cognition (Porsteinsson et al. 2021). But, as with the inevitability of death and taxes, so goes cognitive decline: across almost all measures of cognition, people experience a decline in cognitive ability in old age (see Murman 2015; Salthouse 2010; Tucker-Drob 2019 for reviews). However, the level and rate of decline across race and gender, as I previously reviewed, is modifiable.

Cognitive Reserve and Education: A Mechanism for Cognitive Disparities Across Race and Gender

In order to conceptualize the variation and disparities in cognition in old age, I anchor the concept of cognition and cognitive decline in the theory of cognitive reserve. I find cognitive reserve the most sociologically-relevant theory in the psychology literature to help explain differential rates of ADRDs and cognitive decline across socially defined groups. Researchers theorize cognitive reserve is the adaptability—capacity, efficiency, and flexibility—of the brain to stave off ADRDs and cognitive decline in old age by creating and maintaining denser neurological pathways in the brain (Stern 2002, 2012; Stern et al. 2020). The concept of cognitive reserve provides the neurological underpinnings of why two people with similar levels of brain aging and pathology may have vastly different cognitive abilities (Stern 2002, 2012;

Stern et al. 2020). Cognitive reserve is rarely measured directly because of the need to perform brain scans, and thus is often proxied using educational attainment, which is thought to either improve or reflect the experiences that contribute to cognitive reserve (Stern et al. 2020). Cognitive reserve is thus inextricably linked to education and has the potential to elucidate racial and gender differences in cognition and ADRDs.

Cognitive reserve provides a mechanism for how education translates into cognition: education provides the means for building cognitive reserve, while differential educational environments shapes who has access to cognitive reserve-building experiences (Berkman and Glymour 2006; Glymour and Manly 2008; Mungas et al. 2018; Stern et al. 2020). In the United States, past de jure and current de facto discrimination and segregation highly structured (and structures) educational attainment for people of color and women (Berkman and Glymour 2006; Glymour and Manly 2008). Especially for those who were school-aged before the Civil Rights Act of 1964, race and gender—and their intersection—dictate the level and quality of schooling one receives. In fact, evidence finds that Jim Crow era school segregation (Peterson et al. 2021) and education quality (Liu et al. 2022; Sisco et al. 2015) explains a large portion of the race and gender differences in cognition, while others find that increased high school graduation rates from 1900 to 1950 lowered rates of dementia for both the Black and White population (Hayward et al. 2021). These findings suggest that access to education and education quality influence later life cognition through cognitive reserve. However, these studies evaluate race and gender disparities in cognition under the assumption—without demonstration—of measurement invariance.

Measurement Invariance at the Intersection of Race and Gender

Measurement invariance [detailed further below] is an important—and necessary—characteristic that latent variables must exhibit in order to make cross-group comparisons. It is the assumption that the same underlying factor is captured by the same set of items across social dimensions, like race, gender, and age (Meredith 1993; Meredith and Teresi 2006). In the words of Horn and McArdle, measurement invariance tells us "...whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (1992: 117). This is especially important in the social sciences, where predictors and outcome variables are frequently latent factors—like sentiments, political leanings, or mental health status—that are measured with an index of observed items (Leitgöb et al. 2023; Meredith 1993). These latent factors are then compared across social categories—race, gender, socioeconomic status—to gauge the association between group membership and the latent variable (Leitgöb et al. 2023). However, previous research comparing cognition between socially defined groups often operates under the assumption, and not confirmation, of measurement invariance despite its necessity to make comparisons across social categories (Widaman and Reise 1997).

The substantive importance of measurement invariance is that the average differences between groups may be biased due to language barriers, culturally-specific definitions of words, or differential social desirability (Sass 2011). Researchers risk misspecifying or misinterpreting average differences between groups as "real" differences, when the differences could be attributed to measurement non-invariance. In terms of latent variable analysis with cognition, the previously outlined studies on ADRDs do not explicitly assess measurement invariance, and thus

some proportion of the differences in cognition may be due to measurement non-invariance, suggesting that the disparities we see may not be as extreme as they appear.

There is a small set of studies that assess measurement invariance of cognition across race and gender using samples of older Americans. To measure cognition, researchers often use a neuropsychological assessment of memory and problem solving to gauge cognition, which often asks respondents to perform tasks like memorizing words, solving abstract problems, or naming objects (see Tucker-Drob 2019). In their measurement invariance analysis of the Health and Retirement Study's (HRS) cognition assessment, Blankson and McArdle found that cognition was invariant across gender (measured across men and women), race (measured as non-Hispanic Black, Hispanic, and non-Hispanic White), and time (measured as survey wave) (2015). In other words, the authors found that the latent construct of cognition is captured equivalently across gender, race, and time in the HRS, and thus researchers can directly compare mean estimates across race and gender categories and over time. However, Blankson and McArdle's study did not consider measurement invariance at the intersection of race and gender, and potentially missed bias that is only observable when simultaneously considering race and gender.

One paper investigated measurement invariance of cognition at the intersection of race and gender, in a sample of adults over 65 from northern Manhattan. The authors found evidence for measurement equivalence between all six comparison groups among Black, White, and Latine men and women when particular items were removed from the cognition factor that were not contributing to the measure of Latine women's cognition (Avila et al. 2020). As the authors point out, had the analysis assessed measurement invariance by gender and race/ethnicity separately, they would not have uncovered the particular bias that Latine women experience in their cognitive assessment (Avila et al. 2020). Measurement invariance analysis at the

intersection of race and gender proves to be invaluable for identifying bias in cognition assessments that would have otherwise gone unnoticed.

Intersectionality as a Framework for Measurement Invariance Analysis

Bias due to measurement non-invariance viewed through an intersectional lens offers a framework to understand how previous evaluations of cognition may be misspecified. As Kimberlé Crenshaw originally conceived it, intersectionality is a Black feminist theory outlining the unique vulnerabilities Black women experience with discrimination and violence in the United States (Crenshaw 1989, 2017). Crenshaw established the theory of intersectionality using three legal cases in which Black women brough suit against their employers for discrimination in promotion practices (Crenshaw 1989). Management denied any wrongdoing, referencing their track record of promoting Black men and White women, treating race and gender separately. The courts agreed: Black women were no different from White women in terms of gender and no different from Black men in terms of race (Carbado 2013; Crenshaw 1989, 2017). Crenshaw argued that the court must recognize that both statuses exist simultaneously and acknowledge that Black women should be a separate class of plaintiffs, thus race and gender must be considered simultaneously (1989).

While acknowledging Crenshaw's original impetus was to highlight systems of oppression and erasure of Black women's experiences, intersectionality theory widened the scope of how social scientists view power and inequality (Bauer 2014; Carbado 2013). As a more generalized theory of power relationships, social scientists understand systems of oppression (and privilege) as operating not on a single axis but at the intersection of a multitude of statuses (Bauer 2014). Intersectionality has thus been adopted as a framework for understanding how

social dimensions intersect to create differential exposures to structural inequality that shape a number of outcomes, including health (Cho, Crenshaw, and McCall 2013). For example, researchers have made several calls for intersectional approaches in the fields of psychology (Bowleg 2008, 2012), population health (Agénor 2020), and more generally in quantitative social science research (see Bauer et al. 2021 and Harari and Lee 2021 for reviews), while researchers have utilized intersectionality to evaluate mental health (Ross et al. 2016) and medical care access and treatment (Agénor et al. 2015).

For the purposes of this dissertation, I adopt an intersectional framework to better understand the size and direction of bias in the measurement of cognition in the Health and Retirement Study. This dissertation fits into larger historical and current efforts to accurately measure cognition and diagnose dementia across diverse populations within the United States and internationally. In recent decades, the United States has invested heavily in enumerating the prevalence of dementia in aging Americans. Of particular interest in this dissertation is the Health and Retirement Study (HRS), which has, since its first survey wave in 1992, gathered detailed demographic and cognition data every other year on a probability sample of approximately 20,000 American households where at least one person is aged 51 years or older (Ofstedal et al. 2005; Sonnega et al. 2014). Born out of the HRS was the Aging, Demographics, and Memory Study (ADAMS), which was designed to calculate not only the prevalence of dementia, but additionally the prevalence of pre-clinical cognitive impairment without dementia (cognitive impairment, no dementia, or CIND) (Langa et al. 2005, 2020). The ADAMS was a time- and capital-intensive endeavor, requiring a 3- to 4-hour assessment with a nurse and neuropsychologist to administer the protocol to the 850 HRS respondents aged 70 years and older chosen randomly from the HRS sample (Heeringa et al. 2009; Langa et al. 2005). Though

costly in terms of money and time, the data gathered from ADAMS was invaluable for understanding the prevalence and rates of ADRDs and CIND, which had previously never been assessed. The TICS and ADAMS are important given the rapidly aging population and the potential to understand the mechanisms that lead to the onset of ADRDs.

Outside of the Western context, populations in low- and middle-income countries of the Global South are also rapidly aging. While the projection of the number of people over 60 living in higher income countries is forecasted to increase by 56% between now and 2050, the population of people over the age of 60 is expected to grow by 185% (from 230 to 660 million people) in middle-income countries and 239% (from 33 to 111 million people) in low-income countries (Prince et al. 2015). In conjunction with the rapidly aging global population, there is a growing public health concern surrounding the increased prevalence and future projections of ADRDs in middle- and low-income countries. For example, current estimates place dementia prevalence at 58% in countries the World Banks classifies as low- and middle-income; that percent is expected to grow to 68% by 2050 (Nichols et al. 2022; Prince et al. 2015).

Similar to the U.S. context, the growing global public health concern around ADRDs—the major cause of cognitive disability in aging populations—has led to international initiatives that gather population data to better understanding the etiology of dementia in middle- and low-income country contexts (Gross et al. 2023; Langa et al. 2020). Using the methods, data, and lessons learned from the ADAMS, the HRS established the Harmonized Cognitive Assessment Protocol (HCAP) as an international resource to make cross-national comparisons of the causes, consequences, and trends in dementia around the world (Gross et al. 2023; Langa et al. 2005, 2020). The HCAP has been developed and implemented in international partner studies of the HRS, which include six countries (China, England, India, Mexico, South Africa, and U.S.) with

the intention of providing valid and accurate cognition assessments in diverse cultural, educational, social, economic, and political contexts (Gross et al. 2023; Langa et al. 2020). As the ADAMS was able to enumerate the prevalence of ADRDs and CIND, the HCAP has been able to do the same at a global level, allowing for cross-national comparisons of dementia and pre-clinical cognitive impairment (Gross et al. 2023). While the following analyses focus on the U.S. context, my findings add to the overall undertaking of critically assessing how well these widely-used cognitive assessments capture cognition.

Measurement Invariance

Measurement invariance is an essential, but often overlooked, characteristic of latent variables: it is the assumption that the latent factor of interest—in this case, cognition—is measured equivalently across cultures, geographic space, and time. Measurement invariance is most often performed using multigroup confirmatory factor analysis (Horn and McArdle 1992; Meredith 1993) within a structural equation modeling framework (SEM; Bollen 1989). The general factor model is given by $\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Theta}$, where $\boldsymbol{\Sigma}$ represents the covariance matrix of the observed variables in the model; $\boldsymbol{\Lambda}$ is the matrix of factor loadings (λ) measuring the strength of the relationship between the latent factor variables, ξ, and the vector of observed variables, Y; $\boldsymbol{\Phi}$ is the covariance matrix of the factors, ξ; and $\boldsymbol{\Theta}$ is the covariance matrix of measurement errors for Y (Bollen 1989; Svetina, Rutkowski, and Rutkowski 2020).

The observed variables' means are represented in the equation: $E(Y) = E(\nu + \boldsymbol{\Lambda\xi} + \epsilon)$, where ν is the mean structure of the latent factor or what is essentially the intercept of a linear equation (Svetina et al. 2020). With the assumptions that $E(\epsilon) = 0$ and $E(\boldsymbol{\xi}) = 0$, the equation simplifies to $E(Y) = E(\nu)$. The model is thus easily generalized to multiple populations, so that

each population and group has their own covariance and mean structures, i.e., $\boldsymbol{\Sigma}^{(g)}$ with $\nu^{(g)}$, where $g = 1, 2, \ldots G$ and G is the total number of groups in the analysis (Svetina et al. 2020). All models are estimated using a maximum likelihood approach and their fit to the data is assessed using a χ^2 statistic. The null hypothesis in measurement invariance is that the covariance matrix of each group is the same across all groups, or $H_0: \Sigma^1 = \Sigma^2 = \cdots = \Sigma^G$. In order to test for this, researchers distinguish between four different levels of MI, following a hierarchical structure from least restrictive to most restrictive models: *configural*, *weak* or *metric*, *strong* or *scalar*, and *strict* invariance.

Before running the measurement invariance analysis, as done in the prior studies (see McArdle et al. 2013, Blankson and McArdle 2007, Kline 2015), I first test whether the one-factor model fits the full analytic sample and by single intersectional sub-group. This initial step is testing what is referred to as *dimensional invariance*, which ensures that the measurement model fits the data for each group separately before running the multi-group analyses. Next, I perform the first level of invariance analysis by confirming if the data demonstrates *configural invariance*. Substantively, testing for configural invariance requires that, across all groups, the same parameters (observed variables) are related to the same underlying factor (unobserved variable); each group will thus have its own group-specific factor loadings and intercepts (Horn and McArdle 1992). The configural level of invariance is the most basic level of invariance and must hold before pursuing further invariance testing (Horn and McArdle 1992; Meredith and Teresi 2006).

I determine how well the configural model fits the data by assessing the model fit statistics, specifically the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and the χ^2 statistic. I focus on the RMSEA and CFI as indicators of

model fit because they are superior indices for detecting non-invariance in complex models with large sample sizes (Cheung and Rensvold 2002). Substantively, the RMSEA evaluates the degree to which the estimated model differs from that of a theoretical fully saturated model, where all variance and covariance is explained, thus numbers closer to 0 indicate better fit. The CFI, on the other hand, measures the discrepancy between the estimated model and a theoretical baseline model where none of the variance and covariance is explained, thus a number closer to 1 indicates a better fit (Lai and Green 2016). To employ these model fit statistics, I utilize common cutoffs that previous experts have recommended, specifically applying threshold values at or below 0.05 for the RMSEA and values at or above 0.95 for the CFI to indicate a well-fitted model (Svetina et al. 2020; Wilson et al. 2023). The χ^2 is a general fit index, but is heavily influenced by sample size and may show statistically significant differences in model fit when in fact the model fit differences are negligible (Chen 2007; Cheung and Rensvold 2002).

The next level of invariance testing is assessing *weak* or *metric invariance*, which requires that the factor loadings of the observed items measuring the strength of the relationship with the latent variable are equivalent across groups (Meredith 1993). The null hypothesis is thus H_0: $\Lambda^1 = \Lambda^2 = \cdots = \Lambda^G$ (Svetina et al. 2020). To test for this, one must hold all the coefficients (λ) constant across groups by setting the coefficient to be equal to the reference group (Svetina et al. 2020). Again, researchers use model fit indices to test if the model fits the data worse or the same. In this case, one can test whether metric invariance holds by comparing the model fit between the configural invariance model and the metric invariance model by calculating the ΔCFI, ΔRMSEA, and $\Delta\chi^2$ (Chen 2007; Cheung and Rensvold 2002). Because the χ^2 is sensitive to large sample sizes, and thus might indicate a statistical difference when one does not substantively exist, the literature points to using the ΔCFI and ΔRMSEA (Cheung and Rensvold

2002). To determine if the model fit changed significantly, values of ΔCFI less than -0.005 or ΔRMSEA values greater than 0.010 indicate poorer fit and evidence that invariance at that level does not hold (Chen 2007; Cheung and Rensvold 2002; Svetina et al. 2020). If the fit statistics do not significantly deteriorate, then the next step is to test for the next most restrictive model. At this point, after confirming metric invariance, researchers can formally compare estimated factor variances and covariances over groups, but researchers are advised not to compare the latent means between each group (Kline 2015; Widaman and Reise 1997).

The next level of invariance is *strong* or *scalar invariance*, which requires that, in addition to the factor loadings, the item intercepts are the same across groups. When the latent variable is equal to zero, scalar invariance requires that the means of the observed items are equivalent across sub-group (Horn and McArdle 1992; Widaman and Reise 1997). The null hypothesis is H_0: $\Lambda^1 = \Lambda^2 = \cdots = \Lambda^G$, $\nu^1 = \nu^2 = \cdots = \nu^G$ where the coefficient matrices (Λ^G) and intercepts (ν^G) are the same across groups, set to the values of the reference group (Svetina et al. 2020). Again, researchers use the ΔCFI and ΔRMSEA to determine if the measurement model meets scalar invariance by simply subtracting the CFI and RMSEA of the metric model from the scalar model. If the model fit deteriorates significantly, then the model does not meet the criteria for scalar invariance; if the model fit statistics are unchanged, then the model meets the criteria for configural invariance. Meeting the criteria for scalar invariance is important because this is the level of group-level equivalence necessary to compare mean values of the latent variable of interest; without ensuring scalar invariance, the comparison may not reflect true differences in the construct of interest (Horn and McArdle 1992; Kline 2015).

However, there is a solution to measurement non-invariance at the scalar level with *partial scalar* measurement invariance (Horn and McArdle 1992), where a few items may only

meet the threshold for metric invariance and the majority of items meet the criteria for scalar invariance. Using the example of a simple linear regression, partial scalar invariance is the equivalent of allowing each group to have their own intercept or mean value but only for the items that do not meet scalar measurement invariance (Horn and McArdle 1992). By running models that allow each item to be estimated freely (i.e., allowing group-specific intercepts), we can calculate the model goodness-of-fit statistics in each partial scalar model and compare them to the scalar model, thus calculating the ΔCFI, ΔRMSEA, and $\Delta\chi^2$ (Chen 2007; Cheung and Rensvold 2002). If the model fit statistics improve when particular intercepts are freely estimated—i.e., the ΔCFI and ΔRMSEA meet some threshold—then one can conclude that a partial scalar model better characterizes the underlying cognition factor across the intersectional race-gender groups (Cheung and Rensvold 2002). Importantly, group-specific mean values can be calculated using MGCFA, from which analysts can determine the level of bias simply by calculating unit change between the scalar and final partial scalar model.

HYPOTHESES

Using intersectionality theory, I aim to assess whether cognition is equivalently captured across race-gender intersectional groups—between Black men, Black women, White men, and White women—and, if not, the amount of bias that measurement non-invariance introduces. In my intersectional analyses of measurement invariance of cognition in the Health and Retirement Study (HRS) across Black women, Black men, White women, and White men, I hypothesize the following:

Hypothesis 1: The cognition assessment in the HRS will exhibit scalar measurement non-invariance across intersectional race-gender group (i.e., meaning that average cognition scores

cannot be compared across race-gender group). Substantively, the scalar cognition model will not adequately capture cognition the same way across the intersectional groups. For example, White women with the same level of cognition as White men might not be able to do arithmetic in their heads as well as White men due to socialization around who is encouraged to excel in mathematics and exposure to doing math in one's head. In more technical language, the factor loadings will be equivalent across groups—reflective of a less restrictive model—but the item intercepts will not be the same across group—reflecting that the models do not meet scalar invariance. In statistical notation, the null and alternative hypothesis are the following:

$$H_0: \Lambda^{WM} = \Lambda^{WW} = \Lambda^{BM} = \Lambda^{BW}, \nu^{WM} = \nu^{WW} = \nu^{BM} = \nu^{BW}$$

$$H_A: \Lambda^{WM} = \Lambda^{WW} = \Lambda^{BM} = \Lambda^{BW}, \nu^{WM} \neq \nu^{WW} \neq \nu^{BM} \neq \nu^{BW}$$

where Λ^{WM}, Λ^{WW}, Λ^{BM}, and Λ^{BW} represent the group-specific matrices of factor loadings for White men, White women, Black men, and Black women, respectively, which are equal across groups and in both the null and alternative hypothesis. The vectors of group-specific intercepts are represented by, ν^{WM}, ν^{WW}, ν^{BM}, and, ν^{BW} for White men, White women, Black men, and Black women, respectively, and are equal in the null hypothesis but are all unequal across all race-gender groups in the alternative hypothesis.

Hypothesis 2: Because cognition assessments are positively biased towards White men, Black women will have the largest bias in their cognition assessment due to multiple, intersecting statuses that result in multiplicatively biased cognition assessments that inherently privilege White men. White women and Black men will have lower levels of bias than Black women relative to White men. In statistical terms, the null hypothesis is that the differences between the mean estimate from the scalar and partial scalar models will be zero across groups, while the alternative hypothesis states that the mean difference ($\kappa_{G\ diff}$) will be largest for Black

women, with White women, then Black men following. White men will not have a difference in means because they serve as the reference category, and thus have a difference of zero.

$$H_0: \kappa_{BW\ diff} = \kappa_{WW\ diff} = \kappa_{BM\ diff} = \kappa_{WM\ diff} = 0$$

$$H_A: \kappa_{BW\ diff} > \kappa_{WW\ diff} > \kappa_{BM\ diff} > \kappa_{WM\ diff}$$

METHODS

Data

The Health and Retirement Study (HRS) is a nationally representative longitudinal study of Americans aged over 50. The HRS is an ideal data set for these analyses because of its focus on aging populations and its rich, detailed data collection. Respondents are surveyed once every two years and a new cohort is introduced every six years. The study draws its respondents using a probability sample of households in which at least one member is over 50 years old and non-institutionalized. The survey began in 1992, and has recently released its 15th wave of data for 2020. The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The present study uses data from 1996 (Wave 3—from which point cognition questions are asked consistently) to 2018 (Wave 14), the most recent wave for which sampling weights have been calculated (Health and Retirement Study 2022; RAND Corporation 2022). The sampling weights, produced by RAND Corporation, account for the HRS's sampling structure, which oversamples Black populations, Hispanics/Latines, and Florida residents (Health and Retirement Study 2022; RAND Corporation 2022). RAND additionally harmonizes, cleans, and codes HRS variables into a user-friendly dataset that is easily accessed online (RAND Corporation 2022).

Dependent Variable

Cognition in the HRS is assessed using the HRS-Telephone Interview for Cognitive Status (HRS-TICS). The twelve tests in the HRS-TICS measures two umbrella concepts: episodic memory or mental status. The TICS uses two questions to measure episodic memory: immediate word recall (IMRC) and delayed word recall (DLRC) test. Respondents are given a list of ten words from four possible lists and are asked to immediately recall as many words from the list to the best of their ability (IMRC) and are asked five minutes later to do the same (DLRC) (McArdle et al. 2007). Scores are between zero and ten for both tests.

The remaining ten tests measure mental status, and include: serial 7s test (count back from 100 by 7 for 5 trials; scored 0-5); counting backwards from 20 for ten continuous digits in two trials if needed (scored 0-2); identifying a cactus and a pair of scissors after listening to a description (scored 0 or 1 for each); correctly naming the current president and vice president (scored 0 or 1 for each); and stating the day's date (day, month, year, day of the week; scored 0-4). Summing together the memory and mental status scores, respondents have total possible cognition scores from 0 to 35 (McArdle et al. 2007; McCammon et al. 2022; Ofstedal et al. 2005).

For the purposes of this study, the cactus, scissors, president, and vice president dummy variables are recoded into a single four-point ordinal categorical variables (referred to as "Names"). The same is done with the date dummy variables, and are referred to it as "Dates". The recoding is necessary because these two sets of dummy variables are highly correlated and fit the data better as two ordered categorical variables as opposed to eight dummy variables. Lastly, the counting backwards from 20 variable is recoded to a dummy variable, where respondents who successfully completed the test in one or two trials are assigned "1" (named

"CNTB"). I recoded this because very few people were unable to count backwards from 20 on their first try, and thus would not contribute to the measurement model of cognition. More substantively, I wanted to differentiate between who was able versus unable to complete the task, and not necessarily to what degree respondents were able to complete the task. In addition, these variables are recoded as I have described in other foundational works on measurement properties of the HRS's cognitive function factor (McArdle et al. 2007; Ofstedal et al. 2005).

An important difference between the following analyses and previous measurement invariance analyses with HRS data is that I utilize a single-factor measurement model as opposed to a two-factor model. The two-factor model has memory (immediate and delayed word recall) and mental status (Ser7, CNTB, Names, Dates) factors under the umbrella factor of cognition (Blankson and McArdle 2015; McArdle 2011; McArdle et al. 2007). The one factor model simply puts all six items into a single factor. I chose to use the one-factor model to more easily interpret the results, but have supplemental analyses with the two-factor model that substantively reflect similar results as the one-factor model (see Supplementary Table 2 and Supplementary Figure 1)

Predictor Variables

The intersectional identities of particular interest in this chapter are respondents' race, ethnicity, and sex (referred to as gender hereafter). Race is measured as Black, White, or other in the HRS. I exclude in this study respondents who identified as other, leaving only those who self-identified as Black or White. I made this choice because the Black and White categories make up the majority of racial identities in the HRS and, more importantly, the heterogenous

"other" category does not have racial distinctions in the Public Use data; if I were to include it in my analysis, I would not know which racial(ized) category respondents belong to.

Ethnicity is measured as self-identifying as Hispanic or non-Hispanic. I excluded those who identified as Hispanic because of the heterogeneity in Hispanic origin, which is only crudely categorized as "Mexican" and "other" in the Public Use data. While this decision is imprecise, I made it on the basis that people who identify as non-Hispanic White and non-Hispanic Black are less likely to speak English as a second language. If I were to introduce Hispanic/Latine respondents, there is a strong likelihood that the source of measurement non-invariance would be due to language barriers, as opposed to focusing on bias due to the hegemony of White men. A crosstab of the race and ethnicity variables resulted in the two race/ethnicity groups of interest in this study: non-Hispanic White (referred to as White hereafter) and non-Hispanic Black (Black hereafter).

Gender was measured as a binary, either female. I acknowledge that conceptualizing gender beyond a simple binary is a worthwhile investigation in terms of measurement invariance and is well established in gender studies as an intersectional social status that brings with it systems of oppression. However, besides the fact that the HRS only provides "male" and "female" answer options, my theoretical motivation for this paper relies primarily on the existence of (White) male hegemony. Future studies could include measurement bias related to non-binary respondents, but is a facet that goes beyond the purview of this chapter. The intersecting social statuses of interest in this chapter are race/ethnicity (Black and White) and gender, which combine to create four sub-groups: Black men, Black women, White women, and White men. These race-gender intersectional groups represent the social statuses that either (in

the case of White men) benefit from or (in the case of the former three groups) are harmed by the primacy of White patriarchy and its ingress in cognition assessments.

Current Study Sample

The starting sample in the HRS for Waves 3 (collected in 1996) to Wave 14 (collected in 2018) is N = 232,326 individual observations among 39,958 respondents. In order to be included in the analytic sample, respondents had to have non-missing values for race, ethnicity, gender, and sampling weights. I also only included those aged 50 years and older and only included the first time a respondent answered the TICS sequence, as McArdle et al. 2007 did for their cross-sectional measurement invariance analysis. I use these sample restrictions because the HRS is a representative sample of American's aged over 50, and thus people enter the study and take the cognition assessment for the first time starting at age 50. Additionally, I included the first time people took the TICS to control for any serial correlation and because there is evidence that respondents superficially improve their cognition score due to prior test experience, which would further bias the results (Salthouse 2010, 2019). The final sample size is N =30,576 respondents, with 10,523 White men, 13,240 White women, 2,748 Black men, and 4,065 Black women.

Summary sample statistics include unweighted means and standard deviations of respondent's age, birth cohort, average years of schooling (Table 1) and across the six cognition items (Table 2) for the full sample and across race-gender intersectional group. All summary and descriptive statistics were conducted using Stata 17 (Stata Corp 2021). The average age for the full sample was 62.6 years, with a standard deviation of 10.0. Ages ranged from 50 to 105. Average age by intersectional group appears to be more similar by race category than by gender. For instance, Black men and women are approximately 60 years old at their initial testing while

White men and women are, on average, 62 to 64 years old at their first assessment. Similarly, the average birth cohort for Black men and women is around 1945 compared to 1937 to 1938 for White men and women. Lastly, years of education hover around 12-13 years, except for Black men who have 11.7 years of education. Group differences shown in these summary statistics illustrate that there is a potential for cognition to be additionally invariant by age and/or education, and not just by race-gender intersectional group. Age varies widely across the sample, and might be another factor to consider for measurement invariance, but is outside the scope of this chapter. Education, on the other hand, is fairly stable across race-gender intersectional groups, and is unlikely to be an underlying factor for measurement non-invariance.

Table 2 presents descriptive statistics of the six cognition items for the full sample and by intersectional race-gender group. Overall, it appears women outperform men within each race category for tasks related to memory (IMRC and DLRC). However, mean differences are fairly mixed for remaining four variables. Men perform better than women within race group on the serial 7s task, while the discrepancy in counting backwards appears to be down racial lines, where White men and women are more successful in completing the task than Black men and women. Differences in dates appears to be similar for Black women and White men, highest for White women, and lowest for Black men. Lastly, discrepancies in names appear to be down racial lines, with Black men and women answering closer to 3.2-3.3 of the items correctly compared to an average of 3.7 for White men and women.

Figure 1 presents the average overall cognition score, scored from 0 to 35, across the race-gender intersectional groups. Overall, the scores are fairly similar, but have very small confidence intervals, and thus are all statistically different from one another. Ultimately, these descriptive statistics show the largest group difference appears along racial lines: White men and

women have slightly higher cognition scores compared to Black men and women. The differences between Black men and women, as well as White men and women, is not substantively large. As is, these are the overall average cognition scores we would expect without performing a measurement invariance analysis. Without ensuring that the cognition factor demonstrate measurement invariance, these average overall scores may be biased, leading to problematic interpretations of "real" cognitive abilities.

Analytic Plan

I conducted a multigroup confirmatory factor analysis (MGCFA) within a structural equation modeling (SEM) framework to test for measurement invariance of cognition by intersectional race-gender groups. Substantive measurement invariance analyses were conducted using Mplus version 8.8 (Muthén and Muthén 2017). I do this by estimating a series of one-factor models using weighted least square with mean and variance adjusted estimation (WLSMV), each with increasingly strict parameter restrictions, measured across the four intersectional sub-groups. Theta parametrization was employed to identify the model, which sets residual variances to one (Millsap and Yun-Tein 2004; Paek et al. 2018).

There are many ways to identify MGCFA models depending on one's research question. The mean values that are produced from the MGCFA models will be the basis for calculating the proportion of bias measurement non-invariance introduces in the mean estimate of cognition, and thus are imperative to calculate. The measurement model I chose to analyze is a one-factor model, where all observed items contribute directly to the cognition factor. The one-factor model departs from that of McArdle et al. (2007) and Blankson and McArdle (2015), all of whom used a two-factor model. After running both the one- and two-factor models, I found that the models

do not substantively differ in model fit, and thus chose the one-factor model for ease of interpretation.

Before running MGCFA, the model must be properly identified—as is, the model cannot be identified because there are more parameters than there are unique values to predict them (Kline 2015). In order to identify the model, I must scale the latent variable, since it does not have an inherent metric. I do this by fixing the factor loadings for IMRC to one, which sets the scale of the factor to the metric of the IMRC variable. I made this decision on the basis that the reference variable should not be chosen arbitrarily, but based on its non-invariance i.e., the "most non-invariant" item (Cheung and Rensvold 1999), which can only be achieved after determining which items are noninvariant after estimating the model (Sass 2011). The remaining coefficients are freely estimated, meaning they are not set to a particular number and represent—in the scale of the IMRC variable—the strength of the relationship between each item and cognition. In effect, each intersectional group has their own group-specific coefficients. Additionally, I hold the IMRC intercept to be equal to the reference group's value (i.e., White men) across the sub-groups. In my analyses, the reference group is White men because, in my conceptualization of intersectionality, White men hold the most power and privilege. Deviations from White men—the benefactors of both privilege and the absence of gender- and race-based discrimination—will indicate deviations from White patriarchal constructs of cognition reflected in the testing instrument.

After identifying the models, I run four increasingly restrictive models: configural, metric, scalar, and partial scalar models. The configural model, which is the most basic level of invariance, reveals how well the latent variable structure fits the data of each sub-group. To fit this model, I restrict the model structure and freely estimate the latent variable factor loadings,

intercepts and thresholds, and mean cognition score across intersectional group (save for IMRC factor loadings and intercepts, which are set to that of White men). The next most restrictive model is the metric model, which requires that the factor loadings of the observed items measuring the strength of the relationship with the latent variable are equivalent across groups (Meredith 1993). To test for this, I set the factor loadings to be equal to the reference group, White men, across the groups. Next, the scalar invariance model restricts the factor loadings and item intercepts (or thresholds for categorical variables) to be the same as the reference group across intersectional group. After running the scalar invariance model, I run six separate models in which each item is separately estimated freely (i.e., each intersectional group has their own group-specific intercept or threshold). For the sake of brevity, I do not interpret each of the six items in the paper, but provide the output in Supplementary Table 1.

To determine whether model fit deteriorates or improves across the configural, metric, scalar, and partial scalar models, I utilize three goodness-of-fit indices: the χ^2, the root mean square error of approximation (RMSEA), and the Comparative Fit Index (CFI). The χ^2 is a general fit index, but is heavily influenced by sample size and may show statistically significant differences in model fit when in fact the model fit differences are negligible (Chen 2007; Cheung and Rensvold 2002). As Cheung and Rensvold (2002) recommend, I focus on the RMSEA and CFI as indicators of model fit because they are superior indices for detecting non-invariance in complex models with large sample sizes. To gauge overall model fit, an RMSEA value smaller than 0.05 and CFI values greater than 0.95 indicate good model fit (Svetina et al. 2020; Wilson et al. 2023). To measure whether model fit improves significantly from each model to the next more restrictive model, I utilize the ΔRMSEA and ΔCFI, which are simply the differences between the indices from two models. To determine if the model fit changed significantly, values

of ΔCFI less than -0.005 or ΔRMSEA values greater than 0.010 indicate poorer fit and evidence that invariance at that level does not hold (Chen 2007; Cheung and Rensvold 2002; Svetina et al. 2020). I found that Ser7, Names, and Dates, when freely estimated (see Supplementary Table 1 for full output), contributed the most to improving model fit. These three items are freely estimated in the final partial scalar model, which, in effect, is the "new" scalar model. I compare the final partial scalar model to the metric model.

The last set of analyses will use information gathered from the measurement invariance analyses to calculate a set of means and standard deviations that correct for any measurement non-invariance by sub-group. The scalar model represents what the mean value would by group if each item had the same latent factor loadings and intercepts, which is the default in latent variable analysis or regression analysis. Each item must have the same average relationship across each group. The subsequent mean value estimated from the partial scalar models allows each group to have their own group-specific thresholds, removing parts of the measurement model that are introducing bias into the average. To ascertain the amount of bias, I subtract the mean estimates from the scalar model from the mean estimates from the partial scalar model, in effect calculating the amount of bias that measurement non-invariance introduces.

RESULTS

Single-Group Confirmatory Factor Analysis

Before moving into the MGCFA, as done in the literature (see McArdle et al. 2013, Blankson and McArdle 2007, Kline 2015), I first tested whether the one-factor model fits the full analytic sample and by single intersectional sub-group. This initial step ensures that the model fits each intersectional group's data; however, this does not provide much substantiative

information, other than that all the cognition items are measuring the factor. Across the full sample and across each sub-group, the RMSEA is below 0.05 and the CFI is above 0.950 indicating that the models fit each group's data fairly well. The data indicates that dimensional invariance has been achieved, which simply means that the model was able to run and fit reasonably well. This is the baseline level of invariance necessary to move forward with measurement invariance testing. Dimensional invariance indicates that the six items are related to the cognition factor across the four intersectional groups.

Multi-Group Confirmatory Factor Analysis

The results of the configural model (Table 4) reflect what the cognition means would be if each sub-group was able to have its own group-specific coefficients, intercepts, and error variances. The model fit indices indicate a well-fitted model: the RMSEA is 0.016 and CFI is 0.996. The mean values across the intersectional groups do not follow the gradient that I had initially expected in regards to the intersectionality theory. White women appear to outperform White men, while Black women perform slightly worse than White men, and Black men rank lowest in cognitive test scores. Next, the metric model, where the factor loadings are set to that of the White male group across the intersectional groups, appear to fit the data better than the configural model: the RMSEA is smaller than that of the configural model and the CFI is the same. The mean values in this case do not change in the metric model compared to the configural model.

The scalar model—or the model that would produce latent mean cognition values in a exploratory factor analysis—sets the factor loadings and intercepts/thresholds to be the same as White men. The model has a precipitous drop in model fit, with a RMSEA of 0.911 and CFI of

0.044. Here, the RMSEA falls below the threshold of 0.95, providing evidence that the scalar model does not fit the data. Additionally, the ΔRMSEA is 0.03 and ΔCFI is -0.085, indicating the model fit deteriorated significantly between the metric and scalar models. Importantly, the mean values of cognition in the scalar model mimic that of the overall average cognition score in the summary statistics presented above. White women and men have similar average values, with White women having a slightly negative average value. Similarly, Black women and men have closer mean values, but have significantly lower mean values compared to White women and men.

In Figure 2, on the lefthand side, I have plotted the mean estimates from the scalar model for each intersectional sub-group. Visually, it is clear that the mean values for White women and men and Black women and men are close together. In terms of Hypothesis 1, I find evidence to support the alternative hypothesis: although the factor loadings are similar across intersectional group, the intercepts and thresholds are not the same across intersectional group. These results suggest a partial scalar model is necessary.

In the partial scalar model, the thresholds for the Ser7s, Dates, and Names tests were freely estimated (i.e., each intersectional group had its own group-specific set of thresholds for Ser7s, Dates, and Names). Previous analyses (see Supplementary Table 1) suggested that these three variables contributed the most to the scalar model not fitting the data well; when allowed to vary by intersectional group, the partial scalar model fits the data much better. The RMSEA drops to 0.016, with a ΔRMSEA of 0.002 between the partial scalar and metric models; both indices indicate that the final model meets the previously defined thresholds for a well-fitting model. The CFI is 0.994, with a ΔCFI of -0.002; again, these two metrics indicate the partial scalar model does not deteriorate significantly in fit from the metric model.

The estimated averages from the partial scalar model increase from the scalar model across the three intersectional groups, nearing the estimates from the configural and metric models. Black women's mean cognition estimate increases to -0.433, reflecting a 0.483 unit increase between the scalar and partial scalar models. White women's average increases as well, changing to a positive value well above White men at 0.369, with a 0.439 unit increase in the estimated mean. Black men's cognition estimates show a more modest increase to -0.777, indicating a 0.144-unit increase.

In terms of Hypothesis 2, I do not find evidence for the null or alternative hypothesis; instead, there are more substantial increases for Black and White women, while Black men saw a modest increase in their mean value. Figure 2 displays the partial scalar mean estimates on the righthand side. The difference between the scalar and partial scalar models is very clear: Black women and White women outperform their male counterparts. The unit change in the mean values from the scalar to partial scalar model reflects the amount of bias that the scalar model introduces. By allowing each intersectional group to have their own group-specific intercepts for Ser7s, Dates, and Names, the gap between Black men and women and White men narrows significantly, while White women have significantly higher cognition averages compared to White men.

DISCUSSION

Previous research has established that cognition in aging Americans follows an expected race- and gender-gradient reflective of historical and structural discrimination and marginalization: Black people's cognition fairs worse than that of White people, while women bear the brunt of Alzheimer's in old age compared to men. Researchers have hypothesized that

the relationship between historical and current racism and sexism has influenced education access and quality, leading to lower levels of cognitive reserve—a characteristic of aging brains that can prevent or slow the onset of cognitive decline. However, previous studies on race- and gender-based disparities in cognition have largely neglected the potential for bias in the cognition instruments themselves. A potential answer lies in measurement invariance analysis: using a MGCFA within an SEM framework, analysts can pinpoint if and to what extent a latent variable is measured equivalently across socially defined groups. From there, researchers can evaluate the amount of bias that measurement non-invariance introduces.

Bias due to measurement non-invariance viewed through an intersectional lens offers a framework to understand how previous evaluations of cognition may have been misspecified. My analyses show that the mean estimates from the scalar model reflect a cognition gradient that falls along racial lines: Black men and women have similar estimates, which are below those of White men and women. However, in the partial scalar model, where each intersectional group has their own group-specific intercepts, the model fit improves, while the mean estimates increase across the groups. For one, Black women saw a 0.483 unit increase in their estimated cognition score, while White women's estimate increases by 0.439 units. Essentially, when I remove the bias due to measurement non-invariance, Black and White women's cognition estimates grow, outperforming their male counterparts. Black men saw a more modest increase, with their mean estimate growing by 0.144 units, approximately one-fourth of the increase that Black and White women saw.

While I find evidence for Hypothesis 1, where cognition does not meet the threshold for scalar invariance, I did not find evidence to support Hypothesis 2. In Figure 2, the estimates very clearly diverge from the scalar to partial scalar model, where the estimates increase significantly

for White and Black women, both of whom have higher average estimates compared to their male counterparts. These findings differ from those of Blankson and McArdle (2015), who did not find differences by race and gender, separately. In the one-factor model, I do find variation at the intersection of race and gender, which would have otherwise gone unnoticed had I not considered the intersection of race and gender. Notably, I find substantively similar results when I run these analyses with a two-factor model: the memory factor is invariant (i.e., equivalently measured across race-gender groups) while the average mental status estimates increases for Black women and White women when bias is removed (Supplementary Table 2 and Supplementary Figure 1).

While correcting for the bias that measurement non-invariance introduces to the latent cognition variable narrows the gap between Black women and men and White men, a large and significant gaps still exists across all four intersectional groups. The measurement invariance analysis has provided important information on how much bias exists, but does not completely close the gap. The gap might be due to measurement invariance due to age or education, where the items in the cognition variable operate differently as people age or reflect cognition differently depending on respondents' level of education, a topic for future studies. Aside from measurement invariance per se, cognition may be influenced by age and education outside of measurement bias. Ultimately, these findings open the potential for the design of a cognition assessment developed through qualitative methodology and grounded theory. Perhaps through these methodologies, there is the potential to develop a culturally- and gender-informed cognition assessment that critically evaluates a universal measure of cognition.

As with any study, these analyses come with some limitations. At the extreme end of possible limitations is that the six test items in the HRS do not measure global cognition but

measure some other latent factor related to cognition—this is a single factor model and may not capture the fullness of a person's cognitive ability. The measurement invariance analysis does not necessarily negate this possibility; however, these items have been tested and validated across multiple populations with painstaking care, and thus are unlikely to be completely unrelated to cognition (Herzog and Wallace 1997; McArdle et al. 2007). A less extreme limitation is controlling the sample for age-related measurement non-invariance. As noted in the summary statistics of the sample, White women tended to be older than Black women in the sample. Just as testing for cognition might vary by race and gender group, measurement invariance may not hold across time and age. For example, the ability to count backwards from 100 by sevens may indicate functional cognition among people aged 65 and younger, but is not a test reflective of functional cognition or cognitive decline among those aged over 65. This limitation opens the door for future directions that utilizes these partial scalar invariance and models how these mean values of mental status and episodic memory change over time.

Despite these limitations, this study finds evidence for measurement non-invariance and substantial bias associated with it. Without viewing these mean differences by intersectional race-gender groups, these average disparities in cognition would go unnoticed. Importantly, these analyses add to the effort of the HRS and its international partners to better understand the intricacies of measurement equivalence across diverse populations and how perhaps small differences in measurement equivalence translate into disparities in cognition estimates.

TABLES AND FIGURES

Table 1. Unweighted descriptive statistics of the Health and Retirement Study (HRS) 1996-2018, N=30,576

		Age	Year of birth	Years of education
Total	Mean	62.6	1939.4	12.7
N=30,576	SD	10.0	15.2	2.9
100.0%	Min	50	1890	0
	Max	105	1965	17
Black men	Mean	59.7	1945.7	11.7
n=2,748	SD	8.5	14.2	3.4
9.0%	Min	50	1900	0
	Max	100	1965	17
Black women	Mean	59.8	1944.9	12.7
n= 4,065	SD	9.1	14.9	2.9
13.3%	Min	50	1892	0
	Max	103	1965	17
White men	Mean	62.9	1938.1	13.1
n=10,523	SD	10.0	14.5	3.0
34.4%	Min	50	1897	0
	Max	98	1965	17
White women	Mean	63.7	1937.3	12.8
n=13,240	SD	10.3	15.2	2.6
43.3%	Min	50	1890	0
	Max	105	1965	17

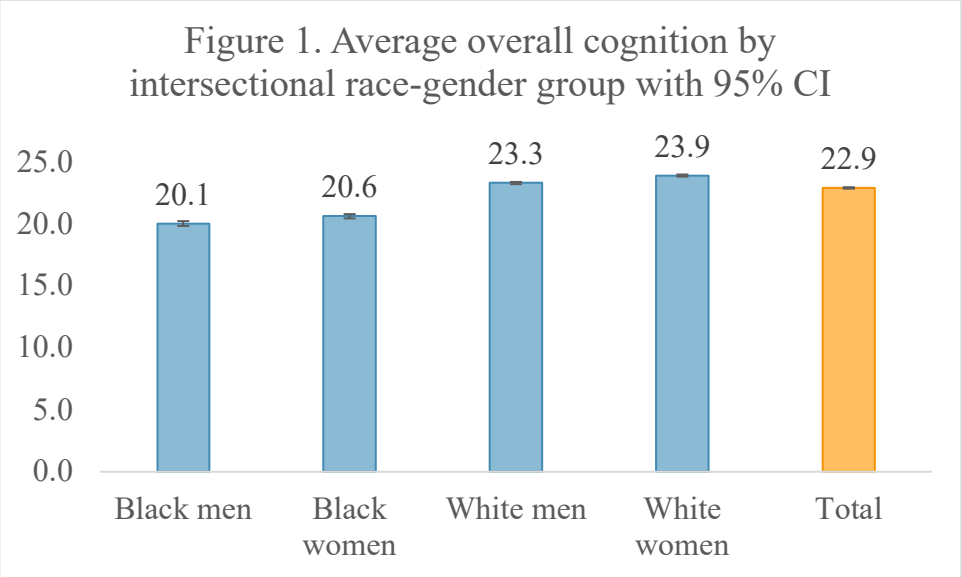


Table 2. Mean and (standard deviation) of cognition items by intersectional race-gender group, Health and Retirement Study (HRS) 1996-2018, N=30,576

	IMRC	DLRC	Ser7	CNTB	Dates	Names
Total sample	5.65	4.49	3.56	0.95	3.78	3.62
	(1.73)	(2.11)	(1.65)	(0.21)	(0.53)	(0.65)
Black men	4.93	3.50	2.81	0.90	3.67	3.34
	(1.64)	(1.90)	(1.79)	(0.30)	(0.66)	(0.81)
Black women	5.36	4.01	2.53	0.90	3.77	3.24
	(1.71)	(2.11)	(1.82)	(0.29)	(0.57)	(0.84)
White men	5.54	4.38	4.00	0.97	3.76	3.75
	(1.70)	(2.00)	(1.40)	(0.18)	(0.54)	(0.54)
White women	5.97	4.93	3.68	0.97	3.83	3.68
	(1.71)	(2.12)	(1.55)	(0.17)	(0.46)	(0.57)

IMRC = Immediate word recall; DLRC = Delayed word recall; CNTB = Count backwards from 20; Ser7 = Serial 7s; Dates (day, month, year, day of the week); Names = Naming objects/public figures

Table 3. Single-group Confirmatory Factor Analysis by Intersectional Sub-Group, Health and Retirement Study (HRS) 1996-2018, N=30,576

	Chi-square	df	p-val	RMSEA	95% CI		CFI
Full sample	47.841	7	<0.001	0.013	0.009	0.016	0.997
Black men	20.286	7	0.005	0.026	0.013	0.040	0.991
Black women	10.059	7	0.185	0.010	0.000	0.023	0.999
White men	29.828	7	<0.001	0.018	0.011	0.024	0.996
White women	23.865	7	0.001	0.013	0.008	0.020	0.997

Table 4. Measurement invariance mean estimates across model by intersectional race-gender group and fit statistics by model, Health and Retirement Study (HRS) 1996-2018, N=30,576

	Configural (1)	Metric (2)	Scalar (3)	Partial Scalar: Ser7, Dates, Names (4)	Unit change: Mean(4)-Mean(3)
Cognition					
Black men	-0.691***	-0.687***	-0.921***	-0.777***	0.144
S.E.	0.054	0.054	0.042	0.048	
Black women	-0.328***	-0.323***	-0.916***	-0.433***	0.483
S.E.	0.046	0.046	0.041	0.042	
White women	0.371***	0.375***	-0.070**	0.369***	0.483
S.E.	0.034	0.034	0.025	0.031	
Model fit statistics					
Chi-square	84.119	103.406	1406.572	140.639	
df	28	43	88	49	
RMSEA	0.016	0.014	0.044	0.016	
ΔRMSEA	—	0.002	0.025	0.002	
CFI	0.996	0.996	0.911	0.994	
ΔCFI	—	0.000	-0.085	-0.002	

Note: Ser7 = Serial 7s; Dates = Dates (day, month, year, day of the week); Names = Naming objects/public figures

***p <0.001 **p<0.01 *p<0.05.

Configural invariance (1): factor loadings and intercepts/thresholds are estimated freely so that each group has their own group-specific factor loadings and intercepts/thresholds; Metric model (2): factor loadings are set to that of the reference group (White men), while the intercepts/thresholds are freely estimated; Scalar model (3): factor loadings and intercepts/thresholds are set to that of the reference group; Partial scalar model (4): factor loadings and the intercepts/thresholds for IMRC, DLRC, and CNTB are set to that of the reference group, while Ser7s, Dates, and Names are freely estimated.

The mean estimate is quantified as the sub-group compared to the reference group (White men), which is set to zero.

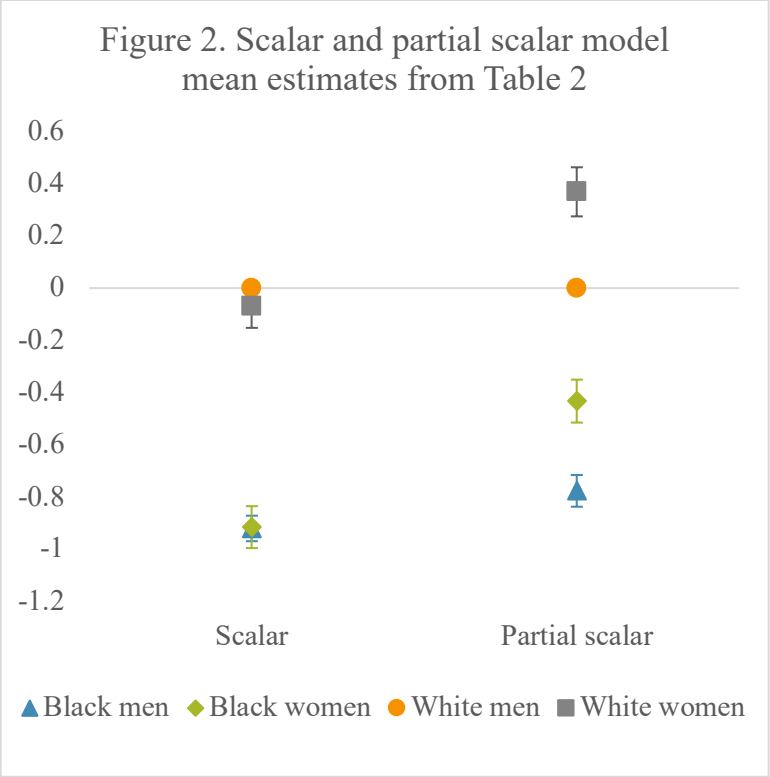


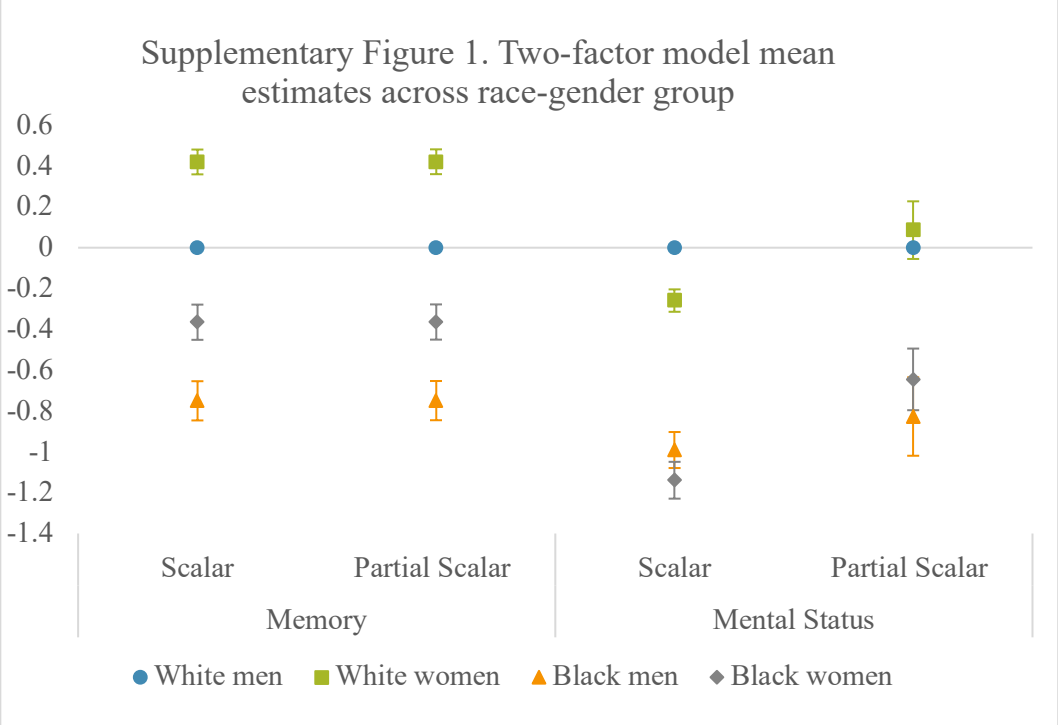
Figure 2. Scalar and partial scalar model mean estimates from Table 2
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
Scalar
Partial scalar
Black men
Black women
White men
White women

SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1. Measurement invariance mean estimates across model by intersectional race-gender group and fit statistics by model in the one-factor model, Health and Retirement Study (HRS) 1996-2018, N=30,576

	Config. (1)	Metric (2)	Scalar (3)	Scalar, IMRC (4)	Scalar, DLRC (5)	Scalar, CNTB (6)	Scalar, Ser7 (7)	Scalar, Dates (8)	Scalar, Names (9)	Partial scalar (10)	Change: Ave.(10) -Ave.(9)
Cognition											
BM	-0.691***	-0.687***	-0.921***	-1.018***	-0.937***	-0.917***	-0.804***	-0.945***	-0.921***	-0.777***	0.144
S.E.	0.054	0.054	0.042	0.046	0.042	0.042	0.044	0.041	0.043	0.048	
BW	-0.328***	-0.323***	-0.916***	-1.073***	-0.992***	-0.921***	-0.596***	-0.966***	-0.885***	-0.433***	0.483
S.E.	0.046	0.046	0.041	0.046	0.042	0.041	0.043	0.040	0.041	0.042	
WW	0.371***	0.375***	-0.070**	-0.140***	-0.141***	-0.074**	0.209***	-0.099***	-0.054*	0.369***	0.439
S.E.	0.034	0.034	0.025	0.027	0.026	0.025	0.027	0.025	0.026	0.031	
Model fit statistics											
Chi-sq.	84.12	103.41	1406.57	1121.87	1106.89	1406.07	655.16	1147.01	1336.67	140.64	
df	28	43	88	85	85	85	73	76	76	49	
RMSEA	0.016	0.014	0.044	0.040	0.040	0.045	0.032	0.043	0.047	0.016	
ΔRMSEA	—	-0.002	0.025	0.026	0.026	0.031	0.018	0.029	0.033	0.002	
CFI	0.996	0.996	0.911	0.930	0.931	0.910	0.961	0.927	0.915	0.994	
ΔCFI	—	0.000	-0.085	-0.066	-0.065	-0.086	-0.035	-0.069	-0.081	-0.002	

Note: BM = Black men; BW = Black women; WW = White women.
Ser7 = Serial 7s; Dates = Dates (day, month, year, day of the week); Names = Naming objects/public figures
***p <0.001 **p<0.01 *p<0.05.
Configural invariance (1): factor loadings and intercepts/thresholds are estimated freely so that each group has their own group-specific factor loadings and intercepts/thresholds; Metric model (2): factor loadings are set to that of the reference group (White men), while the intercepts/thresholds are freely estimated; Scalar model (3): factor loadings and intercepts/thresholds are set to that of the reference group; Scalar, IMRC (4)–Scalar, Names (9): factor loadings and intercepts/thresholds except for the item indicated are set to that of the reference category, while the indicated item is freely estimated; Partial scalar model (10): factor loadings and the intercepts/thresholds for IMRC, DLRC, and CNTB are set to that of the reference group, while Ser7s, Dates, and Names are freely estimated.
The mean estimate is quantified as the sub-group compared to the reference group (White men), which is set to zero.


Supplementary Figure 1. Two-factor model mean estimates across race-gender group
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
Scalar
Partial Scalar
Scalar
Partial Scalar
Memory
Mental Status
White men
White women
Black men
Black women

CHAPTER 3. SECOND-ORDER LATENT GROWTH MODEL OF COGNITION AT THE INTERSECTION OF RACE AND GENDER: REMOVING BIAS REDUCES DISPARITY BETWEEN BLACK WOMEN AND WHITE MEN

ABSTRACT

In the broader social sciences and specific field of demography, scholars have made a concerted effort to understand and explain differences in cognition trajectories by race and gender in old age. Many studies on cognitive decline attempt to explain race and gender differences using sociologically-relevant contextual variables—such as education, socioeconomic status, stress, and geography—that moderate the association between race, gender, and cognitive decline. However, these commonly explored individual and contextual variables fail to fully account for race- and gender-based disparities potentially for reasons rooted in how cognition is conceptualized and measured. A potential answer lies in assessing longitudinal measurement invariance—which is the foundational assertion that latent variables are captured equivalently over social dimensions and time—in cognition over not just race and gender, but at their intersection. Using data from 1996 to 2018 in the Health and Retirement Study (HRS), I calculate a second-order latent growth model, which permits me to calculate the amount of bias in the measurement of cognition over age. Removing bias from the cognition age-trajectories of Black women, Black men, White women, and White men results in a 50% reduction in the gap between Black women and White men. This novel modeling strategy demonstrates the importance of assessing and correcting for bias in the measurement of cognition trajectories across race-gender groups.

INTRODUCTION

In the broader social sciences and specific field of demography, scholars have made a concerted effort to understand and explain differences in cognition trajectories by race and gender in old age. The endeavor is for good reason: as the U.S. population continues to age and become more diverse (U.S. Census Bureau 2018; Vespa et al. 2020), the prevalence of cognitive impairment is expected to grow by 60% between 2020 and 2050 (Alzheimer's Association 2020), with the burden of Alzheimer's and related dementias (ADRDs) falling disproportionately on the shoulders of more Black people than White people, and more women than men (Rajan et al. 2021; U.S. Census Bureau 2021). These enumerations reflect a pressing public health concern. Social scientists frequently utilize large, nationally-representative, longitudinal surveys that collect data on cognition to explain these race- and gender-based disparities by controlling for social and contextual factors that systematically differ by race and gender. However, the ability of commonly explored individual and contextual variables to fully account for such disparities—such as education (Farina et al. 2020; Hale, Schneider, Mehta, et al. 2020; Hayward et al. 2021), socioeconomic status (SES) (Faul et al. 2021; Fujishiro et al. 2019), stress (Chen et al. 2022; Cintron et al. 2023), and geography (Byrne and Anaraky 2022; Pohl et al. 2021)—are limited. This may be for reasons rooted in how cognition is conceptualized and measured. Thus, a question remains: what accounts for the race and gender differences in cognition trajectories over age? A potential answer lies in longitudinal measurement invariance over race and gender.

Social scientists often calculate latent variables in order to compare differences across social dimensions and time. Measurement invariance is a foundational assertion that the latent variable of interest—here, cognition—is captured by the same set of questionnaire items over social dimensions and time (Horn and McArdle 1992). Without ensuring that latent variables are

invariant (i.e., that the latent variables are measuring the same underlying construct across social groups and time), it may be that scholars are comparing proverbial apples to oranges, and not apples to apples (Clark and Donnellan 2021; Meredith and Teresi 2006). Potentially, some proportion of the differences in cognition in the aforementioned studies are not real differences in cognitive abilities but an artifact of measurement non-invariance, where survey items and tests do not capture the same underlying construct across social dimensions and age (Horn and McArdle 1992). Much like other "universal" measures of intelligence or aptitude, previous analyses find that race and social class are the best predictors of success on IQ tests and the SATs (Wicherts 2016; Wicherts and Dolan 2010; Zwick 2019). Cognition assessments may suffer from a similar flaw: cognition questions may not capture a universal measure of cognition, but measure the ways that societal expectations and norms shape people's cognitive abilities. For example, cognition assessments frequently include a memorization task. Due to gendered socialization and expectations around memory—like remembering birthdays, engagements, and names—women perform better on this task than men. This item would be considered biased if women systematically performed better on this task than men, despite both groups having similar levels of cognition. It is thus imperative to test and adjust for measurement non-invariance in age trajectories of cognition over race and gender because the bias due to measurement non-invariance may be interpreted as "true" race, gender, and age differences in cognition.

Bias in the measure of cognition may incorrectly lead researchers to conclude that there are differences in cognition over race and gender when, in reality, the disparity is much smaller. In order to ensure unbiased measurements of cognition trajectories over race and gender, I utilize a type of growth model: a second order latent growth model, discussed in detail in the Background section. This strategy permits me to compare how group-specific cognition

trajectories behave in a biased, unadjusted model relative to an unbiased, adjusted model that modifies the cognition measurement. This approach ensures that cognition is invariant (i.e., measured equivalently) across race and gender groups. Not only is measurement invariance by race and gender important separately, but it must be considered at their intersection to fully account for the complexities that multiple, intertwining social statuses have on cognition that may be missed when statuses are treated separately. In addition, I borrow from demographic research on lifespan variation (Firebaugh et al. 2014; Sasson 2016) to conceptualize and hypothesize how variation in cognition trajectories behave across intersectional race-gender groups. I use longitudinal data from the Health and Retirement Study (HRS) in order to determine how four intersectional race-gender groups differ in their initial cognition in early old age (starting at age 65) and how the group-specific averages change over time into old age to age 80. By comparing an unadjusted, biased model to an adjusted, unbiased model, I can quantify how much bias measurement non-invariance introduces into the intercept, slope, and variation of the trajectories. The unadjusted model will reflect the findings of previous scholars that have not taken into account invariance at the intersection of race and gender over age trajectories. The adjusted model removes the potential bias that I demonstrate superficially diminishes the cognition trajectories of marginalized groups and bolsters the trajectories of privileged ones. This novel modeling strategy generates unbiased measures of cognition over age and intersectional race-gender group.

BACKGROUND

Aging and Lifespan Variation: A Consideration of Cognition

Cognition is a growing concern in the United States given its rapidly aging and changing population. Projections from the U.S. Census estimate that by 2030, one out of five Americans will be over the age of 65 (Vespa et al. 2020) and by 2034, for the first time ever, there will be more adults over the age of 65 than children under 18 years-old (Vespa et al. 2020). In addition, the race and gender composition of people over 65 is projected to change significantly: census projections predict that over the next 40 years, the elderly population will have a larger proportion of Black and Latine people compared to today and a smaller proportion of White people, while women are expected to continue to outlive men, though the sex ratio gap in older ages is projected to decrease over time (U.S. Census Bureau 2018; Vespa et al. 2020). These demographic changes in the aging population are important to understand given the wide disparities in cognitive decline and ADRDs. The prevalence of ADRDs is projected to grow by 60% between 2020 and 2050, barring any major medical breakthroughs to prevent, slow, or cure ADRDs (Alzheimer's Association 2020). Aging—and its co-occurring changes in cognition—as a process is both sociologically and demographically important because it is yet another arena that is subject to social stratification (Crimmins 2020; Crimmins and Zhang 2019; Qiu and Fratiglioni 2018). Given current disparities in ADRDs by race and gender in the United States (Alzheimer's Association 2022; Power et al. 2021; Rajan et al. 2021), the burden of ADRDs will fall disproportionately on the shoulders of more Black people than White people, and more women than men (Chêne et al. 2015; Hebert et al. 2001; Rajan et al. 2021; U.S. Census Bureau 2021) despite overall improvements in the prevalence and incidence of ADRDs over the last 30 years (Langa 2015; Roehr et al. 2018). However, a concept missing from these previous studies

is the quantification of variation in cognitive decline and its implications for disparities in cognition age trajectories. One way to construct a variation argument is to borrow from demographic theories of lifespan variation to conceptualize how variation in cognition trajectories over intersectional race-gender groups is another form of inequality experienced by marginalized groups.

The concept of the "compression of morbidity" in demography is an important characterization of how trends in the predominant causes and rates of mortality have changed over the last century. James Fries proposed that, as lifespans in the population reach their biological limit and as the primary causes of death shift away from infectious disease and infant and maternal mortality to chronic illnesses of old age, mortality rates will be concentrated towards older ages with less variability, leading to a so-called "rectangularization" of survival curves (1980). Fries believe that accompanying this shift in mortality concentration will be a similar shift in morbidity concentration and rectangularization, where people will live longer lives with fewer years lived with disabilities and chronic illnesses (Fries 1980, 2005). However, the prediction was more prescriptive than descriptive. Not only have lifespans continued to rise with little evidence of reaching a biological ceiling (Christensen et al. 2009; Oeppen and Vaupel 2002), but increased mortality has brought with it increased morbidity: though people are living longer, the onset of disability (e.g., functional loss, cognitive decline) and chronic illness (e.g. heart disease, diabetes, osteoporosis) has not kept up with life expectancy, and thus has led to more years lived with morbidities (Crimmins and Saito 2001; Martin, Schoeni, and Andreski 2010; Schneider and Brody 1983). Additionally, researchers have found that despite increases in life expectancy and individual lifespans (both measures of average years of life lived), the variation around when people die has increased (Bohk-Ewald, Ebeling, and Rau 2017;

Engelman, Canudas-Romo, and Agree 2010; van Raalte and Caswell 2013; Tuljapurkar 2011). So-called lifespan variation captures the degree to which populations and sub-groups are nearing a rectangularized survival curve, where the preponderance of death occurs in a small interval in late life (van Raalte, Sasson, and Martikainen 2018). Greater lifespan variation points to both premature and preventable deaths as well as greater uncertainty in mortality outcomes for individuals and greater heterogeneity in populations. Lifespan variation as a marker of inequality is further evidenced by the greater lifespan dispersion among Blacks compared to Whites and those with lower compared to higher levels of education (Firebaugh et al. 2014; van Raalte and Martikainen 2014; van Raalte et al. 2018; Sasson 2016). Thus, not only are the mean estimates of life expectancy important to gauge health disparities, but lifespan variation must also be considered to quantify the health and well-being of populations. The concept of lifespan variation can be generalized to cognition to better understand how variation in cognition trajectories is another facet of inequality.

Cognition is an important facet of health in the later years of life. Though subtle changes in cognition in old age are expected as early as age 60 (Salthouse 2019), these small declines usually do not bar people from being able to continue to take care of themselves (perhaps with assistance), enjoy leisure activities, and have meaningful social relationships. Though the incidence and prevalence of ADRDs have decreased substantially over the last 20 years (Dufouil et al. 2018; Langa et al. 2008; Leggett et al. 2019), the decline in ADRD rates have not been equally distributed: people with higher levels of education have delayed onset of dementia and lower rates of ADRDs compared to those with lower education, while Black women and men have comparatively earlier diagnoses compared to their White counterparts ((Chen and Zissimopoulos 2018; Crimmins et al. 2018; Farina et al. 2020). Just as the lifespan variability

imparts uncertainty on those in marginalized groups (Bohk-Ewald, Ebeling, and Rau 2017; Tuljapurkar 2011), variation between intersectional race-gender group in cognition trajectories over age represent uncertainty for individuals as well as greater heterogeneity at the group- and population-level in cognitive decline. The wide disparities in diagnoses of ADRDs and disparities in cognitive decline by race and gender additionally imply these rates are intervenable and modifiable. Wide variation in cognitive decline and the prognosis and function of cognition poses challenges to marginalized groups that do not have the privilege of certainty. In order to quantify differences in the level of variation between intersectional race-gender groups, I utilize a statistical method that allows me to both account for measurement bias over race-gender groups and age as well as determine the level of uncertainty, or variation, between each group's cognitive decline.

Compression of Cognition with Individual-Level Longitudinal Data

Studies on lifespan variation necessarily use population-level estimates of age-specific mortality and morbidity rates. However, in the following analyses, I model cognition variation with individual-level longitudinal data from the HRS using a variation of a growth curve model. The obvious difference between growth curve models and estimation of lifespan variation is that growth curves follow individuals over time by estimating mean and covariance structures within a structural equation modeling (SEM) framework (McArdle and Epstein 1987; Meredith and Tisak 1990). Results from growth models convey how the outcome variable changes as a result of time or age by producing latent growth factor intercepts and slopes (Byrne 2011; Kline 2015; Wang and Wang 2012). Growth models have the capability of modeling change over time—with

repeated measures within individuals—at the individual- and group-level, or interindividual and intraindivdual level, respectively (Chou, Bentler, and Pentz 1998; Meredith and Tisak 1990).

Aging and declining cognition are often considered together because of their inextricable link, which makes a latent growth model the most appropriate method to quantify cognition over age. The latent growth model treats repeated measures of the outcome variable as a function of age, implying that the central mechanism for the outcome's change is the aging process (Byrne and Crombie 2003; Chou et al. 1998). The basic growth model can be described in the following equation: $y_{ti} = \eta_{0i} + \eta_{1i}\lambda_t + \varepsilon_{ti}$ where y_{ti} is individual *i*'s observed outcome at time *t*; η_{0i} is the latent growth intercept for individual *i*; η_{1i} is the latent slope factor for individual *i*; $\lambda_t$is the time measure for time *t*; and ε_{ti} is the composite error term that includes random measurement error for individual *i*. This equation summarizes the individual- or person-level trajectory. The latent intercept and slope factors can be further reduced to $\eta_{0i} = \eta_0 + \varsigma_{0i}$, where η_0 is the grand mean of the sample intercept and ς_{0i} is the error term in the intercept factor for individual *i*; and $\eta_{1i} = \eta_1 + \varsigma_{1i}$, where η_1 is the overall sample's average change over time and ς_{1i} is the error term for the slope factor individual, *i*. The reduced form simplifies to $y_{ti} = \eta_0 + \eta_1\lambda_t + (\varsigma_{0i} + \lambda_t\varsigma_{1i} + \varepsilon_{ti})$ (Byrne and Crombie 2003; Meredith and Tisak 1990; Wang and Wang 2012). The resulting model is composed of a fixed component ($\eta_0 + \eta_1\lambda_t$) that measures the outcome variable y_{ti} at a specific timepoint, *t*, and random component ($\varsigma_{0i} + \lambda_t\varsigma_{1i} + \varepsilon_{ti}$) that includes three sources of variation: between-person variation around the intercept factor (ς_{0i}); between-person variation in the slope factor ($\lambda_t\varsigma_{1i}$); and within-individual variation over time (ε_{ti}). Importantly, the output from the above model estimates the covariance between ς_{0i} and ς_{1i}, which conveys the association between the initial outcome and change over time (Wang and Wang 2012).

The covariance between the latent intercept and the slope factors is important component for indicating variation in cognition. A statistically significant coefficient for this statistic conveys whether there is an association between the two statistics, i.e., the two estimates are predictable (Pillinger 2020). For example, a negative and statistically significant covariance between the intercept and slope factors indicate that higher initial values of the outcome variable are associated with negative changes over time, which indicates that the trajectories are converging (Pillinger 2020; Wang and Wang 2012). In terms of cognition variation, this represents a more certain, homogenous trajectory. The absence of a statistically significant covariance indicates that there is no relationship between the intercept and slopes. Regardless of the initial level of the outcome variable, the rate at which the outcome variable changes is not statistically predictable, neither converging or diverging (Pillinger 2020). Thus, people with higher cognition do not predictably decline at similar rates, nor do people with lower levels of cognition. Again, as discussed earlier, cognition universally declines as people age; however, a statistically significant covariance between the intercept and slope indicates greater certainty at the individual-level and less heterogeneity at the group-level (Pillinger 2020). Highlighting the often-overlooked covariance of the intercept and slopes further deepens our understanding of inequality in cognitive decline.

HYPOTHESES: UTILIZING INTERSECTIONAL THEORY TO MEASURE INEQUALITY IN COGNITION TRAJECTORIES

Bias due to measurement non-invariance over age viewed through an intersectional lens offers a framework to examine bias in cognition assessments. Popularized by Kimberlé Crenshaw, intersectionality theory argues that multiple social statuses must be considered as

existing simultaneously to acknowledge the unique intersecting, multiplicative systems of oppression those groups experience (Crenshaw 1989, 2017). While acknowledging Crenshaw's original impetus was to highlight systems of oppression and, particularly, erasure of Black women's experiences of discrimination, intersectionality theory widened the scope of how social scientists view power and inequality, especially in the quantitative social sciences (Bauer 2014; Carbado and Roithmayr 2014; Cho et al. 2013). The following analyses consider two areas within which intersectional inequality may be operating: one is at the measurement level and one at the cognition trajectory level. The measurement level considers how much bias there is in the measure of cognition, which is evaluated by performing a measurement invariance analysis (outlined further in the Methods section). The second area is evaluating the trajectories themselves, specifically the intercepts, slopes, and their covariance. As previously noted, I will be comparing two latent growth models: one unadjusted, biased model and one adjusted, unbiased model. The crux of these analyses is to evaluate the intercept, slope, and intercept-slope covariance for the unadjusted and adjusted model, and to compare the estimates from the two models to quantify the level and direction of bias.

Hypothesis 1: Intercepts

In my previous chapter, I found in the cross-sectional sample of adults over the age of 50 that when the measurement model is not adjusted (i.e., in the scalar model), the mean estimates had greater variation by race than by gender: Black women and men had similar cognition estimates that were than the estimates of White women and men. Using this information, I expect the intercepts of the growth model—or average cognition at age 65-66—will reflect similar variation by race and not gender. In statistical notation:

$$\eta_{0,WM} = \eta_{0,WW} > \eta_{0,BM} = \eta_{0,BM}$$

Where $\eta_{0,WM}$ is the grand mean intercept for White men (WM), $\eta_{0,WW}$ is the grand mean for White women (WW), $\eta_{0,BM}$ is the grand mean for Black men (BM), and $\eta_{0,BW}$ is the grand mean intercept for Black women.

In Chapter 2, when I adjusted the measurement model to equivalently measure cognition across the four race-gender intersectional groups, the mean estimates for Black women and White women shifted up, resulting in significantly higher mean values compared to their respective male counterparts. Thus, I expect in the adjusted model that White women will have the highest intercept, followed by White men, Black women, and Black men. In statistical notation, the hypothesis is as follows:

$$\eta_{0,WW} > \eta_{0,WM} > \eta_{0,BW} > \eta_{0,BM}$$

Hypothesis 2: Slopes

The linear slope on age will be negative across all four groups, where cognition is decreasing with age for both the unadjusted and adjusted models. Based on theories of intersectionality, conceptualizations of the compression of morbidity, and my previous hypothesis of how the intercepts will behave, I hypothesize that in the unadjusted, biased model, White women and White men will have similar slopes that are more negative than Black women and Black men. This is because, as I previously hypothesized with regard to intercepts, White women and White men will have the highest intercepts in the unadjusted, biased model, and thus the capacity to have large, negative slopes. In statistical notation, the hypothesis is as follows:

$$\eta_{1,WM} = \eta_{1,WW} < \eta_{1,BM} = \eta_{1,BM}$$

where $\eta_{1,WM}$ is the grand mean slope for the White men (WM) group; $\eta_{1,WW}$is the grand mean slope for White women (WW), $\eta_{1,BM}$ is the grand mean slope for Black men (BM), and $\eta_{1,BW}$is the grand mean slope for Black women (BW).

In the adjusted model, I expect the slopes will become more negative (decrease) for White women and Black women because their intercepts will presumably increase, and thus have the capacity to decline faster. I anticipate Black men and White men will have the same slopes. This will lead to a similar differentiation of slopes as predicted for the intercepts:

$$\eta_{1,WW} < \eta_{1,WM} < \eta_{1,BW} < \eta_{1,BM}$$

Hypothesis 3: Covariance of slope and intercept

The covariance between the slope and intercept of the unadjusted model will be statistically significant and negative (i.e., higher intercepts are associated with more negative slopes) for White women and White men, but will be not statistically significant for Black women and Black men. The reason for this is that White women and men have greater group-level homogeneity in their cognition trajectories compared to Black women and men, reflecting greater certainty in their trajectories:

$$Cov(\varsigma_{0i,WW}, \varsigma_{1i,WW}) < 0, Cov(\varsigma_{0i,WW}, \varsigma_{1i,WW}) < 0$$

$$Cov(\varsigma_{0i,BM}, \varsigma_{1i,BM}) = 0, Cov(\varsigma_{0i,BW}, \varsigma_{1i,BW}) = 0$$

where $\varsigma_{0i,group}$ is the between-person variation in the intercept and $\varsigma_{1i,group}$ is the between-person variation in the intercept. The covariance between $\varsigma_{0i,group}$ and $\varsigma_{1i,group}$ will be significantly less than zero for White women and White men, while the covariance will be indistinguishable from zero for Black women and Black men.

In the adjusted model, I predict the covariance will be negative and statistically significant for Black women, White women, and White men. I believe Black men will not experience a change between the unadjusted and adjusted models because there is little evidence that adjusting for measurement invariance leads to changes in average cognition (as evidenced by results in Chapter 1). In statistical notation, the hypothesis is as follows:

$$Cov(\varsigma_{0i,WW}, \varsigma_{1i,WW}) < 0, Cov(\varsigma_{0i,WW}, \varsigma_{1i,WW}) < 0, Cov(\varsigma_{0i,BW}, \varsigma_{1i,BW}) < 0$$

$$Cov(\varsigma_{0i,BM}, \varsigma_{1i,BM}) = 0$$

METHODS

Data

The Health and Retirement Study (HRS) is a nationally representative longitudinal study of Americans over age 50. The HRS is an ideal data set for these analyses given its focus on aging populations and rich, detailed data collection. Respondents are surveyed once every two years and a new cohort is introduced every six years. The study draws its respondents using a probability sample of households in which at least one member is over 50 years old and non-institutionalized. The survey began in 1992, and recently released its 15th wave of data for 2020 (although without sampling weights). The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. The present study uses data from 1996 (Wave 3—from which point cognition questions are asked consistently) to 2018 (Wave 14), the most recent wave for which sampling weights have been calculated (Health and Retirement Study 2022; RAND Corporation 2022). The sampling weights, stratum, and clustering variables account for the HRS's sampling structure, which oversamples Black populations, Hispanics/Latines, and Florida residents (Health and Retirement

Study 2022; RAND Corporation 2022). These weights are used in the following the analyses. These data are publicly available online, in a harmonized, cleaned, and coded user-friendly dataset produced by RAND corporation (RAND Corporation 2022).

Dependent Variable

Cognition in the HRS is assessed using the HRS-Telephone Interview for Cognitive Status (HRS-TICS). The twelve tests in the HRS-TICS measures two umbrella concepts: episodic memory and mental status. The TICS uses assessments to measure episodic memory: an immediate word recall (IMRC) and delayed word recall (DLRC) test. Respondents are given a list of ten words from four possible lists and are asked to immediately recall as many words from the list to the best of their ability (IMRC) and to complete the same task five minutes later (DLRC) (McArdle et al. 2007). Scores range from 0 and 10 for both tests.

The remaining cognitive tests measure mental status, and include: serial 7s test (count backward from 100 by 7 for 5 trials; scored 0-5); counting backwards from 20 for ten continuous digits in two trials if needed (scored 0-2); identifying a cactus and a pair of scissors after listening to a description (scored 0 or 1 for each; scored 0-2); correctly naming the current president and vice president (scored 0 or 1 for each; scored 0-2); and stating the day's date (day, month, year, day of the week; scored 0-4). Frequently, scholars sum together the memory and mental status scores to create a composite cognition score with possible values from 0 to 35 (McArdle et al. 2007; McCammon et al. 2022; Ofstedal et al. 2005).

For the purposes of this study, I recode the cactus, scissors, president, and vice president dummy variables into a single four-point ordinal categorical variable (referred to as "Names"). The pursue the same strategy for the date dummy variables, referred to as "Dates". The recoding

is necessary because these two sets of dummy variables are highly correlated and fit the data better as two ordered categorical variables as opposed to eight dummy variables. Lastly, the counting backwards from 20 variable is recoded to a dummy variable, where respondents who successfully completed the test in one or two trials are assigned "1" (named "CNTB"). I recoded this because very few people were unable to count backwards from 20 on their first try, and thus would not contribute to the measurement model of cognition. More substantively, I wanted to differentiate between who was able versus unable to complete the task, and not necessarily to what degree respondents were able to complete the task. Notably, the Names and Dates variables are recoded as described in other foundational works on measurement properties of the HRS's cognitive function factor (McArdle et al. 2007; Ofstedal et al. 2005).

Predictor variables

The intersectional identities of particular interest in this chapter are respondents' race, ethnicity, and sex (referred to as gender hereafter). Race is measured as Black, White, or other in the HRS. I exclude in this study respondents who identified as other, leaving only those who self-identified as Black or White. I made this choice because the Black and White categories make up the majority of racial identities in the HRS and, more importantly, the heterogenous "other" category does not have racial distinctions in the Public Use data; if I were to include it in my analysis, I would not know which racial(ized) category respondents belong to.

Ethnicity is measured as self-identifying as Hispanic or non-Hispanic. I excluded those who identified as Hispanic because of the heterogeneity in Hispanic origin, which is only crudely categorized as "Mexican" and "other" in the Public Use data. I made this decision on the basis that people who identify as non-Hispanic White and non-Hispanic Black are less likely to speak

English as a second language. If I were to introduce Hispanic/Latine respondents, it is likely that the source of measurement non-invariance would be due to language barriers, as opposed to focusing on bias due to the hegemony of White men. A crosstab of the race and ethnicity variables resulted in the two race/ethnicity groups of interest in this study: non-Hispanic White (referred to as White hereafter) and non-Hispanic Black (Black hereafter).

Gender was measured as a binary, either male or female. I acknowledge that conceptualizing gender beyond a simple binary is a worthwhile investigation in terms of measurement invariance and is well established in gender studies as a social status that brings with it its own systems of oppression. However, besides the fact that the HRS only provides "female" and "male" answer options, my theoretical motivation for this paper relies primarily on the health impacts of (White) male hegemony. Future studies ought to include measurement bias related to non-binary or transgender respondents. The focal intersecting social statuses in this chapter are race/ethnicity (Black and White) and gender, which combine to create four sub-groups: Black men, Black women, White women, and White men. These race-gender intersectional groups represent the social statuses that either (in the case of White men) benefit from or (in the case of the remaining three groups) are harmed by the primacy of White patriarchy and its centrality in cognition assessments.

Age is the fundamental axis on which cognition changes and, thus, is the central independent variable with regard to changes in cognition. While this presents a missing data issue—there are far fewer people missing across 12 waves of data compared to 50 age categories—Mplus, by default, uses a full information maximum likelihood (FIML) to estimate the models (Muthén and Muthén 2017). The benefit of FIML is that it produces unbiased estimates and standard errors even with missing data (assuming data is missing at random

[MAR]) by estimating a likelihood function for each individual using all available data (Enders and Bandalos 2001). The HRS conducts a full cognition assessment at respondents' initial survey, then again when the respondent is at least 65, and at every wave thereafter (Ofstedal et al. 2005). Between their initial wave to age 65, the respondents are given a partial battery, which only includes immediate (IMRC) and delayed word recall (DLRC), counting back from 20 (CNTB), and counting back from 100 by 7s (Serial 7s) (Ofstedal et al. 2005). To further minimize the aforementioned missing data issue, the sample is restricted to people aged 65 to 80 to avoid the empty cells for the full battery between age 50 to 65. Finally, because respondents were surveyed once every two years, the data were collapsed into two-year age categories starting with age 65-66 up to age 79-80. However, because the HRS did not necessarily interview respondents within the wave year, respondents often were grouped twice in one category because the interview date was less than two years apart from the previous interview. To ensure that respondents had only one observation per two-year age interval, I recoded age to reflect the age that respondents were in the wave year, not the interview year. This resulted in every respondent being in one two-year age category to maximize the sample size and minimize missing data; however, this led to ~25% of the sample to be in an age category one year higher than their actual age. This is the most logical data format since the shift up in age is presumably random and given that the alternative would require separate age categories which would leave one-year age gaps across a majority of respondents.

Analytic Plan

I conducted a multi-group second order (or multiple indicator) latent growth model (SGM) within a structural equation modeling (SEM) framework to test for measurement

invariance of cognition by intersectional race-gender groups over age. Substantive measurement invariance analyses were conducted using Mplus version 8.8 (Muthén and Muthén 2017). The SGM differs from the better-known first order (or single indicator) latent growth curve model (LGCM) in that there are two parts: the measurement model and the latent growth model. The advantage of this two-step process is that it allows researchers to test for measurement invariance of the latent factor over repeated measures and over groups (Chan 1998; Geiser, Keller, and Lockhart 2013; McArdle 1988). Building from chapter one, I estimate separate measurement models for eight two-year age groups (i.e., 65-66, 67-68, 69-70...79-80), focusing only on the scalar and partial scalar models from my previous chapter. The second step utilizes the latent cognition factors at each time point from the first step and models them in a LGCM over eight two-year age groups. The resulting LGCM estimates a latent intercept factor and latent slope factor for each race-gender group. These latent intercepts and slopes represent the growth trajectories, or estimated underlying growth process that leads to differential trajectories in the latent factor (Kline 2015; McArdle and Epstein 1987; Meredith and Tisak 1990). In this chapter, the intercept represents the average latent cognition factor for each race-gender group at age 65 to 66. The latent slope factor measures the average change in the latent cognition factor over age, measured here in two-year intervals. The LGCMs in this chapter are linear models with continuous outcomes.

Before running an SGM, the measurement model must first be identified since, as is, there are more parameters than there are unique values to predict them (Kline 2015). To identify the model, the cognition factor requires a scale since latent factors lack an inherent metric (Sass 2011). According to experts, the scaling variable should not be random, but based on which item in the measurement model is “most invariant” (Cheung and Rensvold 1999; Sass 2011), which in

this case is the IMRC item. The factor loadings for the IMRC item are set to one across all race-gender groups. The remaining factor loadings for DLRC, CNTB, Ser7, Names, and Dates are time-invariant and set to that of the reference group (White males) across the intersectional race-gender groups. In addition, the intercepts for the IMRC item are set to zero at each time point and across all race-gender groups. The two models of interest—the scalar and partial scalar models—differ in how their remaining intercepts are modeled. In the scalar model, the remaining items (DLRC, CNTB, Ser7, Names, and Dates) time-invariant and are set to that of the White male group for all race-gender groups. In the partial scalar model—the specification of which I determined in Chapter 1—the intercepts remain time-invariant, while the DLRC and CNTB intercepts are set to that of the White male group and the Ser7, Names, and Dates intercepts are freely estimated across the intersectional race-gender groups. The time-invariance of the scalar and partial scalar model is a substantive decision given current evidence finds that the cognition factor is invariant over age and time (Avila et al. 2020; Barnes et al. 2015; Blankson and McArdle 2015). This lack of variance means the relationship between item factor loadings and intercepts with the latent cognition factor do not change over age or time.

As previously noted, the focus of this chapter is on the unadjusted or scalar SGM and adjusted, or partial scalar SGMs. An important facet of SGMs is the model fit, which conveys how well the measurement model and LGCM fit the data across intersectional race-gender group (Ferrer, Balluerka, and Widaman 2008). Similar to the multi-group confirmatory factor analysis of the previous chapter, I determine whether model fit improves between the scalar and partial scalar SGMs models. I utilize three goodness-of-fit indices: the χ^2, the root mean square error of approximation (RMSEA), and the Comparative Fit Index (CFI). The χ^2 is a general fit index that is heavily influenced by sample size and may show statistically significant differences in model

fit when in fact the model fit differences are negligible (Chen 2007; Cheung and Rensvold 2002). As Cheung and Rensvold (2002) recommend, I focus on the RMSEA and CFI as indicators of model fit because they are superior indices for detecting non-invariance in complex models with large sample sizes. Substantively, the RMSEA evaluates the degree to which the estimated model differs from that of a theoretical fully saturated model, where all variance and covariance is explained, thus numbers closer to 0 indicate better fit. The CFI, on the other hand, measures the discrepancy between the estimated model and a theoretical baseline model where none of the variance and covariance is explained. Thus, a number closer to 1 indicates a better fit (Lai and Green 2016). To employ these model fit statistics, I utilize common cutoffs that previous scholars have recommended, specifically applying threshold values at or below 0.05 for the RMSEA and values at or above 0.95 for the CFI to indicate a well-fitted model (Svetina et al. 2020; Wilson et al. 2023). To gauge whether model fit improves between the more restrictive scalar SGM and less restrictive partial scalar SGM, I utilize the ΔRMSEA and ΔCFI, which are the differences between the indices from two models. To determine if the model fit changed significantly, values of ΔCFI less than -0.005 or ΔRMSEA values greater than 0.010 indicate poorer fit and evidence that invariance at that level does not hold (Chen 2007; Cheung and Rensvold 2002; Svetina et al. 2020).

In addition, I compare the latent growth factors—intercepts and slopes—between the scalar and partial scalar models across intersectional race-gender group. Similar to calculating the mean change between the scalar and partial scalar models, I calculate the percent change in the intercepts and slopes and additionally plot the cognition trajectories over the four race-gender groups and eight two-year age intervals. The percent difference of the intercepts and slopes between the scalar and partial scalar model represents the amount of bias that the scalar model

introduces into traditional single-indicator latent growth curve model. The partial scalar model's intercepts and slopes will represent the unbiased race-gender cognition trajectories.

Current Study Sample

The starting sample in the HRS for Waves 3 (collected in 1996) to Wave 14 (collected in 2018) is N = 232,326 individual observations among 39,958 respondents. To be included in the analytic sample, respondents had to have non-missing values for race, ethnicity, gender, and sampling weights; respond to the interview themselves and not by proxy; and be aged 65 to 80. The final sample size is N =16,444 respondents, with 965 Black men, 1,556 Black women, 6,006 White men, and 7,917 White women.

RESULTS

Fit Indices for Multi-Group, Multi-Indicator Latent Growth Models

The fit of the scalar model (Table 3) reflects how well the data fit when the factor loadings and intercepts of the cognition measurement model for the growth model are both time-invariant and set to the reference group, which in this case is the White male group. The model fit indices are mixed. The RMSEA is 0.017, which suggests a well-fitted model, with the statistic falling under the threshold of 0.05. The scalar model's CFI is 0.892, which is well below the threshold for a well-fitting model. In the partial scalar model, the intercepts of the Ser7s, Names, and Dates items are estimated freely across intersectional race-gender groups, but are invariant with regard to time. Similar to the scalar model, the fit indices point to a semi-well-fitted model, with an RMSEA of 0.017 and an RMSEA of 0.901. The change in fit between the scalar and partial scalar models suggest that there is a modest change in fit between the two models, with a

ΔRMSEA of 0.00 and a ΔCFI of -0.009. While the ΔCFI meets the criteria for an improved fit of the model, the CFI value is just on the cusp of the threshold for an "adequate" fit.

Intercept and Slope Growth Factors Over Intersectional Race-Gender Groups

The intercepts and slopes across the intersectional race-gender groups did not change drastically between the unadjusted scalar and adjusted partial scalar models. The intercepts for the unadjusted scalar model followed a pattern expected from the partial scalar model, where White women have the highest estimated mean value at age 65-66 (6.34 [0.03]), followed by White men (5.95 [0.04]), Black women (5.30 [0.05]), and Black men (4.83 [0.07]). Thus, I do not find evidence for Hypothesis 1 for the unadjusted scalar model. These unexpected findings may be due to how the estimates change when modeled longitudinally. The slopes—or the predicted change in cognition every two years of age—in the unadjusted scalar model is similar between Black and White men (-0.14), while White women have a slope of -0.15, and Black women a slope of -0.17. Substantively, the slopes are not different from one another, but do fall in line with the expected shape of the trajectories, where cognition is declining from ages 65-66 to 79-80. These findings do not align with the Hypothesis 2, perhaps due to the unadjusted scalar intercepts conforming to what I expected in the adjusted partial scalar model

In the adjusted partial scalar model, the intercepts and slopes change very little from the unadjusted scalar model. For Black men, their intercept increased to 4.90, conferring a 1% increase, which is not statistically significant from the scalar model given that the 95% confidence intervals overlap. Black women saw the largest change, with an intercept of 5.50 in the partial scalar, leading to a statistically significant increase of 4%. White men saw a drop in their intercept value to 5.79, pointing to a 3% decline in their intercepts, which was statistically

different from the scalar model. Lastly, White women saw no appreciable difference in their intercepts from the scalar to the partial scalar model. Here, the evidence for the adjusted partial scalar model in Hypothesis 1 holds for Black women only, where the adjusted partial scalar model led to an increase in the intercept and a removal of 3% of the bias due to measurement invariance. The slopes in the adjusted partial scalar similarly saw little difference from the unadjusted scalar model. Black men and White women's slopes in the partial scalar model remained unchanged at -0.14 and -0.15, respectively. Black women saw a slight decrease in their slope, from -0.17 to -0.16, suggesting that by allowing the Ser7, Names, and Dates intercepts to vary across race-gender group, the slope factor shifted slightly up towards zero. White men also had a small decrease in their slope, from -0.14 to -0.13. Though the Black women's and White men's changes in slopes indicates a 6% increase, these differences were not statistically significant, evidenced by the overlapping 95% confidence intervals for the scalar and partial scalar estimates for each group. Thus, there is little evidence to support Hypothesis 2 in regard to the adjusted partial scalar model.

A visualization of the trajectories is presented in Figure 1, where the solid lines refer to the partial scalar model and the dashed lines refer to the scalar model. The adjusted partial scalar model allows the intercepts of the Ser7, Names, and Dates items to be freely estimated across each group, creating race-gender-specific intercepts for those three variables. The largest differences between the scalar and partial scalar models appears to be between Black women and White men. Comparing the scalar and partial scalar trajectories, Black women's slope jumps up and while White men's deteriorates, leading to a clear narrowing of the gap between Black women and White men. For Black women, this increase in the intercept and slight decline in the slope leads to a consistently and statistically higher estimated mean cognition value compared to

Black men, whereas in the scalar model, the mean estimates for Black women at ages 77-78 and 79-80 overlap with Black men's estimates. Taken together, these findings suggest that, while White women did not see the hypothesized change in their trajectories, the removal of a small percent of bias in the measurement of cognition led to an attenuated race-gender gap between Black women and White men, where Black women's trajectories shifted upward and White men's shifted down.

The last set of results concerns the covariance between the intercept and slopes. In the unadjusted scalar model, the covariance estimates align with Hypothesis 3: the covariance between the intercepts and slopes for White women and White men are negative and statistically significant. The same statistics are not statistically significant for Black women and Black men. This reflects the variation in intercept and slope, where there is a predictable relationship between the two parameters for White women and White men. Black women and Black men do not have the same predictable relationship, meaning that the slope may be steep or shallow regardless of starting cognition. In the adjusted partial scalar model, the covariance between the slope and intercepts remained unchanged, which did not support the expectations for the partial scalar model in Hypothesis 3. Again, White women and White men had statistically significant negative covariances between the slopes and intercepts, while Black women and Black men had insignificant covariance values. This is most likely due to the minor differences in the intercept and slopes between the scalar and partial scalar models. Thus, despite removing a small amount of bias due to measurement invariance did not change the amount uncertainty and variation that Black women and men experience in regards to their cognition. By allowing each intersectional group to have their own group-specific intercepts for Ser7s, Dates, and Names in their cognition

trajectories over age, the gap between Black women and White men narrows significantly, but does not change the amount of dispersion in trajectories across intersectional race-gender groups.

DISCUSSION

Previous research has established that cognition in aging Americans is discrepant across race and gender, such that Black people's cognition fairs worse than that of White people's, and women bear the brunt of Alzheimer's disease in old age compared to men. In trajectories over age, researchers have attempted to explain the gaps between race and gender groups by controlling for relevant social factors, like education, SES, stress, and geography, but often still find race- and gender-based gaps. However, previous studies on race- and gender-based disparities in cognition trajectories over age have neglected the potential for bias in the cognition assessment themselves. The potential response to this oversight lies in utilizing second-order latent growth curve models (SGM) within a structural equation modeling (SEM) framework, where analysts can assess if and to what extent cognition trajectories are measured equivalently across social dimensions. In addition, these analyses highlight the utility of an often-overlooked growth model statistics: the covariance between the individual variation in the intercept and individual variation in the slope. These estimates inform whether adjusting for measurement non-invariance across groups is associated with less variation, and thus less individual uncertainty and greater population-level heterogeneity, in cognition trajectories. The analyses in this study utilize an intersectional lens to understand how previous evaluations of cognition may have been misspecified. These analyses are not an indictment of the cognition assessment in the HRS. Instead, study results add to the HRS's important work that aims to increase the precision of the cognition assessment across diverse populations globally and across time.

The analyses show that the cognition trajectories in the unadjusted scalar model are similar to the cross-sectional partial scalar model estimated in Chapter 2, where White women have the highest intercept, followed by White men, Black women, and Black men. The estimates do not change drastically in the adjusted partial scalar growth model, where each intersectional group has their own group-specific intercepts for Ser7s, Names, and Dates. However, there is a significant reduction in the disparities between Black women's and White men's cognition trajectories. While the slopes do not change for White men and Black women between the unadjusted scalar and adjusted partial scalar model, the intercepts do: Black women see a statistically significant rise in their cognition intercept, while White men have a statistically significant decrease in their cognition intercept. The result is a clear narrowing of the gap between Black women's and White men's cognition trajectories as these populations age. The implication of these analyses is that utilizing a model that allows Black women and White men to have their own group-specific intercepts in the measurement portion of the SGM narrows the disparity in cognition between Black women and White men.

The findings of the covariance between the intercept and slopes of the models show that, regardless of adjusting for measurement non-invariance, Black women and Black men have greater variation and dispersion in their cognition trajectories, while White women and White men have statistically predictable, converging cognition trajectories. These results suggest that not only are there inequalities across intersectional race-gender groups in terms of average cognition over age, but Black women and men face the additional facet of inequity in the amount of uncertainty in their cognition trajectories. Conversely, White women and men have the privilege of having greater certainty and population homogeneity in cognition. These findings deepen our understanding of the breadth of inequality that exists in cognition.

While adjusting for measurement non-invariance narrowed the gap between Black women and White men, a large and significant gap still exists across all intersectional race-gender groups. Additionally, there is greater uncertainty in the cognition trajectories of Black women and Black men. However, the disparities in cognition trajectories and variation in cognition trajectories may be due to differences in the level of education, exposure to stress, and geographic residence. Though, as previously noted, extant literature has been unable to fully account for race- and gender-based disparities in cognition trajectories, the addition of adjusting for measurement non-invariance in models that control for social context may eliminate discrepancies in cognition trajectories. Additionally, the inclusion of contextual variables may be able to explain some of the variation in trajectories. I investigate the potential mediating influence of these forces in Chapter 4.

As with any study, these analyses come with limitations. For one, the second-order latent growth model was challenging to run, in that the covariance matrix across groups had to strike a precarious balance between the latent factors being highly correlated, but not too correlated. This made it challenging to model a fully configural model, where all parameters (factor loadings and intercepts) are freely estimated across the four groups and eight time points. While there is little evidence in the literature that factor loadings and intercepts vary over age (Avila et al. 2020; Barnes et al. 2015; Blankson and McArdle 2015), there may be an answer to modeling a configural model of this size and complexity by utilizing an alignment-within-CFA (AwC) growth curve model (Asparouhov and Muthén 2014, 2023; Muthén and Asparouhov 2018). This technique utilizes alignment optimization within a CFA framework (Marsh et al. 2018) that allows researchers to perform an unbiased growth curve model that adjusts for measurement non-invariance across many time points and across many groups (Lai 2021). These limitations

open the door for future endeavors to model the complexity inherent in cognition over age, race, and gender. Despite the limitations, this study finds evidence for measurement non-invariance in cognition trajectories over intersectional race-gender groups. Without viewing these trajectories by intersectional race-gender groups over age, these biases in cognition would go unnoticed. Additionally, by utilizing the covariance between the slope and intercept, I find evidence for another characteristic of inequality: variation in cognition trajectories. Study findings suggest that future analyses of differences in cognition across socially defined groups must take into account the potential for measurement bias in cognitive measurement tools when modeling changes over age across race-gender groups.

TABLES AND FIGURES

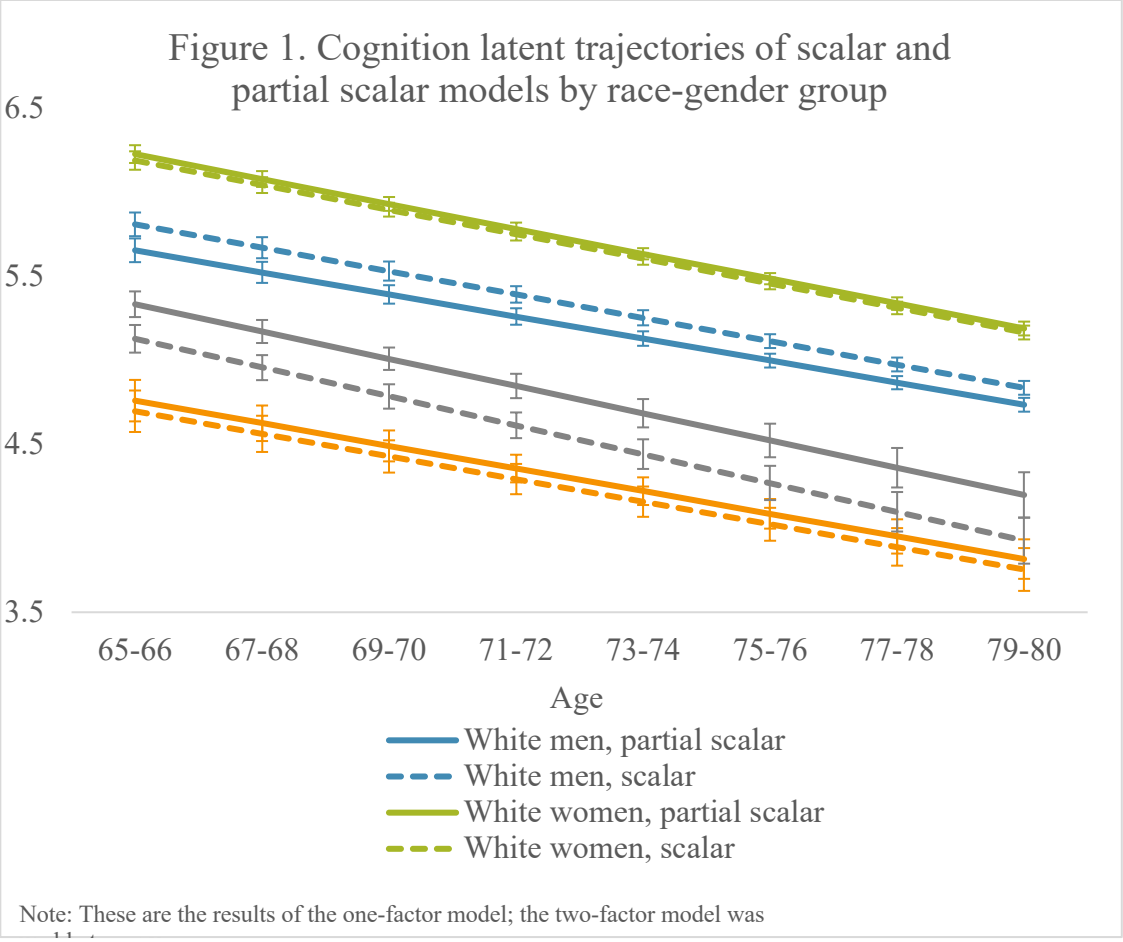
Table 3. Goodness-of-fit indices for the scalar and partial scalar models

	Scalar	Partial scalar
Chi-square	9476.17	9025.24
df	4258	4250
RMSEA	0.017	0.017
ΔRMSEA	—	0.000
CFI	0.892	0.901
ΔCFI	—	-0.009

Table 4. Scalar and partial scalar model growth factor estimates (95% confidence interval) of the intercept and slope by intersectional race-gender group

	Scalar			Partial scalar			% change:		
	Int. (1)	Slope (2)	Cov (Int, Slope) (3)	Int. (4)	Slope (5)	Cov (Int, Slope) (6)	Int. (3) – Int. (1)	Slope (4) – Slope (2)	Cov (3) – Cov (4)
BM	4.83***	-0.14***	0.02	4.90***	-0.14***	0.02	1%	0%	0%
95% CI	(4.97, 4.58)	(-0.16, -0.11)	(-0.01, 0.05)	(5.03, 4.75)	(-0.16, -0.11)	(-0.01, 0.05)			
BW	5.30***	-0.17***	-0.03	5.50***	-0.16***	-0.03	4%	6%	0%
95% CI	(5.39, 5.20)	(-0.19, -0.15)	(-0.08, 0.02)	(5.59, 5.41)	(-0.18, -0.14)	(-0.08, 0.02)			
WM	5.95***	-0.14***	-0.03***	5.79***	-0.13***	-0.03***	-3%	6%	%
95% CI	(6.02, 5.87)	(-0.13, -0.15)	(-0.05, -0.01)	(5.86, 5.71)	(-0.14, -0.12)	(-0.04, -0.01)			
WW	6.34***	-0.15***	-0.03***	6.38***	-0.15***	-0.03***	1%	1%	0%
95% CI	(6.40, 6.27)	(-0.16, -0.14)	(-0.04, 0.02)	(6.43, 6.31)	(-0.16, -0.14)	(-0.04, 0.02)			

Note: BM = Black men; BW = Black women; WM = White men; WW = White women.
The partial scalar model allows the intercepts of Ser7s, Dates, and Names to vary by race-gender group.
Ser7 = Serial 7s; Dates = Dates (day, month, year, day of the week); Names = Naming objects/public figures.



Note: These are the results of the one-factor model; the two-factor model was

CHAPTER 4. LONGITUDINAL MEASUREMENT INVARIANCE AT THE INTERSECTION OF RACE AND GENDER: ADJUSTING FOR CONTEXTUAL FACTORS NEARLY ELIMINATES DISPARITIES FOR BLACK WOMEN, BLACK MEN, AND WHITE MEN

ABSTRACT

In the United States, cognitive decline is subject to stratification along social dimensions. A central goal of health disparities research is to identify the sociological source of these disparities. However, the explanatory power of socially-relevant measures falls short: despite controlling for stressors, education, and geography, disparities by race and gender persists. A question remains: what accounts for these disparities? Most research on race and gender disparities in cognition is executed under the assumption—without demonstration—of measurement invariance. Potentially, some proportion of the differences in cognition do not reflect actual differences in cognitive abilities but are an artifact of measurement non-invariance because survey items and tests do not capture the same underlying construct across sub-groups. The following analyses add to the extant research by adjusting cognition age-trajectories for measurement bias while simultaneously controlling for relevant covariates. Using data from Health and Retirement Study (HRS), I calculate an unbiased second-order latent growth model and control for three measures that proxy the individual (biomarkers of stress), institutional (education), and structural (region) factors that influence cognition. Education had the largest influence on the level, but not change, in cognition across all four intersectional race-gender groups. When I control for all three variables, the disparities in cognition trajectories between Black women, Black men, and White men are substantively eliminated, while White women's cognition remains at a comparatively high level. These analyses illustrate the potential explanatory power of measurement invariance analyses for explaining the remaining gaps in cognition trajectories.

INTRODUCTION

In the United States, cognition and Alzheimer's and related dementias (ADRDs) are subject to stratification along social dimensions. Cognition and ADRD disparities are especially apparent along race and gender lines. Census estimates (Rajan et al. 2021; U.S. Census Bureau 2021), incidence rates (Steenland et al. 2016; Weuve et al. 2018), and estimated lifetime risk (Power et al. 2021) of ADRDs all point to a race-cognition gradient that advantages the White over the Black population. Women, on the other hand, make up nearly two-thirds of ADRD cases in the United States (Rajan et al. 2021), though this may be attributed to women's longevity, and thus greater lifetime risk, and not due to sex per se (Fitzpatrick et al. 2004; Hale, Schneider, Gampe, et al. 2020; Hebert et al. 2001; Levine et al. 2021). These enumerations reflect a pressing public health concern. Social scientists frequently utilize large, nationally-representative, longitudinal surveys that collect data on cognition to explain these race- and gender-based disparities by controlling for social and contextual factors that systematically differ by race and gender. However, the explanatory power of socially-relevant measures falls short: despite controlling for stressors (Chen et al. 2022; Cintron et al. 2023), education (Farina et al. 2020; Hale, Schneider, Gampe, et al. 2020; Hayward et al. 2021), and geography (Byrne and Anaraky 2022; Pohl et al. 2021), disparities by race and gender persists. A question remains: what accounts for these disparities?

Most research on race and gender differences in cognition is executed under the assumption of measurement invariance. Measurement invariance is a foundational assertion that the latent variable of interest—here, cognition—is captured by the same set of tests and questionnaire items over social dimensions and time (Horn and McArdle 1992). However, it may

be that some proportion of the differences in cognition do not reflect actual differences in cognitive abilities but are an artifact of measurement non-invariance because survey items and tests do not capture the same underlying construct across sub-groups (Horn and McArdle 1992). Take, for example, any popular standardized test, like the SATs or IQ tests: these tests are attempting to assess some universal measure of cognitive ability or aptitude, but previous analyses find that race and social class are the best predictors of success on these tests (Wicherts 2016; Wicherts and Dolan 2010; Zwick 2019). The same applies to cognition assessments: it would be problematic to suggest, after looking at the disparities by race and gender, that these differences are due to some inherent population-subgroup deficiency. Instead, the assessment questions may not capture a universal measure of cognition, but reflect gendered and racialized societal norms and expectations. For example, cognition assessment items on a survey may task respondents with memorizing a set of words. Due to lifetime gendered socialization and expectations around memorizing important dates, relationships, and names, women may perform better on this task than men. This item would be considered biased towards women if women systematically performed better on this task than men, despite both groups having similar levels of cognition. Thus, some proportion of the group differences in average cognition may be due to mismeasurement in the latent variable itself.

Extant research has not adjusted cognition trajectories for measurement non-invariance while simultaneously controlling for relevant covariates. The remaining disparities in cognition trajectories are possibly due to measurement non-invariance across race and gender group. Additionally, the disparities due to measurement non-invariance may be compounded at the intersection of race and gender. As separate statuses, cognition trajectory estimates by race and gender may conceal disparities in cognition trajectories that can only be revealed when

considering their intersection. In addition to accounting for measurement invariance, the analyses pursued in this chapter include three covariates of interest: individual-level biomarkers, education, and residence in the southern United States. The contribution of the following analyses is to use the unbiased cognition trajectories (which are adjusted to remove measurement non-invariance) across age and intersectional race-gender groups, and to control for these three measures that proxy the individual (biomarkers), institutional (education), and structural (region) factors that influence cognition. By controlling for the measurement bias in the cognition assessment in the Health and Retirement Study (HRS) and these three factors central to research investigating cognition, I explain the majority of the race-gender inequality in cognition trajectories that has remained largely unresolved.

BACKGROUND

In sociology and demography, disparities in cognition and ADRDs are understood to originate from social structures that create the contexts in which race and gender, and their intersections, are consequential for cognitive health. Facets of social identity, like race and gender, shape cognition trajectories through the prism of historical and current forms of marginalization and structural discrimination. A useful framework for understanding race-gender disparities in cognition is through the social of determinants of health (SDOH), which asserts that health disparities are the result of larger, structural factors that unequally distribute access to resources that prevent, delay, and lessen the severity of illness (Braveman et al. 2010, 2011; Braveman and Gottlieb 2014). The conditions and environments in which people live, learn, work, and age across the lifespan determine resources like employment opportunities, access to education, housing, and health care, all of which are germane to health and cognition outcomes

(Braveman and Gottlieb 2014). Social, economic, political, and historical power structures determine the distribution of these resources, which systematically disenfranchises historically marginalized groups while maintaining the hegemony of the elite (Adkins-Jackson et al. 2023). These conditions are often referred to as "upstream" factors because they are considered the so-called "causes of the causes" (Braveman et al. 2011). These differential exposures may explain some of the disparities in cognition trajectories.

The social determinants of health framework contextualizes how cognition disparities arise from the power structures that unequally distribute exposures and access to health promoting factors. The observed differences in cognition trajectories in old age suggest the existence of modifiable risk factors that can delay the onset and severity of cognitive decline and ADRDs. Viewed through the SDOH framework, modifiable risk factors exist at every social level, including at the individual, biological level; institutional level; and the structural (including geographic) level. Each level is associated with a unique association with cognition and offer potential mechanisms for explaining the differences in cognition trajectories over race-gender group. Based on prior studies, I have chosen allostatic load, education, and years lived in the South as the covariates of interest that are most relevant to cognitive aging especially when considered across race-gender intersectional groups.

Association Between Chronic Stress, Allostatic Load, and Cognition

People encounter and experience stressors every day. The perception of a stressful event activates a chemical barrage that focuses the body's resources and energy to navigate a short-term threat, or a "fight or flight" response (McEwen and Seeman 1999; McEwen and Stellar 1993). While quotidian short-term dips and peaks in stress is normal—referred to in the

psychology literature as allostasis, or when the body's physiological systems are operating within normal bounds (McEwen and Stellar 1993)—chronic and elevated stress responses are biologically costly, increasing the risk of chronic, long-term illness (Diwadkar 2016; Karlamangla et al. 2002; Seplaki et al. 2006). In particular, continued stress responses leads to dysregulation of the metabolic, cardiovascular, and immune systems, leading to greater risk of diabetes and heart disease (McEwen and Seeman 1999; McEwen and Stellar 1993). The resulting excess biological markers (i.e., biomarkers) of stress is called allostatic load, which measures the accumulation of physiological dysregulation and the associated so-called "wear and tear" on the body's biological systems, including cognition (McEwen and Seeman 1999; McEwen and Stellar 1993). The effects of allostatic load on cognition are of particular interest because the brain is directly linked to perceptions of and responses to stress and subsequent stress dysregulation (Bruce S McEwen 2016; McEwen and Gianaros 2010). Studies have found a clear association between higher allostatic load (AL) values and worse global cognition and executive function (see D'Amico, Amestoy, and Fiocco 2020 for review and meta-analysis). Ansell et al. (2012) found that cumulative stress leads to diminished gray matter volume in the brain, which is thought to play a significant role in memory and cognition. Importantly, research finds a clear overlap between parts of the brain that are affected by stress and those that undergo the greatest atrophy in aging (see Ganzel, Morris, and Wethington 2010 for review). Persistent, elevated stress is thus detrimental to cognition especially when conceptualized as a socially structured exposure that can shape disparities in cognition over race and gender.

An important factor to take into account is the unequal distribution of stressors throughout society that disproportionately affect marginalized groups. According to the social determinants of health framework, social structures determine exposure to persistent, chronic

stressors by race and gender, and their intersections, leading to elevated AL and worse cognition in old age for select groups. The United States is stratified by race through past de jure and present de facto racism and discrimination. Compared to White people, Black people are overrepresented in measures of poverty, including income, neighborhood disinvestment, and receipt of public assistance (Bailey et al. 2017; Williams and Collins 2001; Williams and Sternthal 2010). The same structural forces that disenfranchise the Black population disproportionately expose those in poverty to the chronic stressors that unstable housing, precarious employment, and food insecurity brings (Beech et al. 2021; Sternthal, Slopen, and Williams 2011; Williams, Priest, and Anderson 2016). Besides overrepresentation in poverty measures, people of color experience racism at the interpersonal, institutional, and structural level in the United States (Robert Wood Johnson Foundation 2017). Speaking specifically about Black people in the United States, research has found that the experience of racism is detrimental to physical and mental health (Anderson 2013; Williams 1999; Williams and Mohammed 2013) as well as cognition (Letang et al. 2021; Zuelsdorff et al. 2020). The combination of economic disenfranchisement and experiences of racism puts Black people at higher risk for elevated AL, leading to worse cognition.

In addition to the consideration of race, researchers must examine the influence of stress and AL on cognition at the *intersection* of race and gender. Black women in the U.S. experience gendered racism, a specific form of discrimination that is a unique, chronic stressor (Harnois and Ifatunji 2011; Jackson, Hogue, and Phillips 2005; Thomas, Witherspoon, and Speight 2008). Living in a sexist and racist society as a Black woman means bearing the burden of expectations and discrimination that those two statuses bring (Perez et al. 2023). “Double Jeopardy” (Beal 2008) and “Multiple Jeopardy” (King 1988) are early conceptualizations of the multiple forms of

oppression that Black women, who embody historically and currently marginalized statuses, face in American society. Evidence finds that Black women experience discrimination through microaggressions (e.g., slights, invalidation, insults), harassment, prejudice, and stereotyping that is unique to being both Black and a woman (Nuru-Jeter et al. 2009; Sue et al. 2007). This is slightly different from intersectionality, which also acknowledges the multiple intersecting privileged statuses that advantage hegemonic statuses.

In order to navigate and exist in a society that continuously perpetrates this discrimination, Black women must utilize coping mechanisms to remain resilient in the face of these forms of violence in the form of the "Superwoman Schema" (Knighton et al. 2022; Perez et al. 2023; Woods-Giscombé 2010). The Superwoman Schema was developed to better understand the sociohistorical context of gendered racism in the United States, where Black women report the need to present an image of stoicism, strength, resilience, independence, and obligation to others (Woods-Giscombé 2010). While in some ways an adaptive coping mechanism—the Superwoman Schema helps Black woman maneuver gendered racism at the interpersonal, institutional, and structural levels—research to date has found that Black women who identify with the Superwoman Schema have worse health outcomes related to stress and AL, leading to subsequently worse cognitive outcomes (Allen et al. 2019; Coogan et al. 2020; Rodriguez et al. 2019). The additional unique stressors that Black women face is important to take into account when investigating stress, AL, and cognition.

Cognition and Education: The Role of Cognitive Reserve

The link between education and cognition in old age has been thoroughly examined in the social sciences, and continues to be a central factor in cognitive health in aging populations. The

literature has established that the association between education and cognition in aging adults is positive: people with higher levels of education have, on average, higher cognition, while those with lower levels of education have comparatively lower cognition (see Lenehan et al. 2015; Maccora, Peters, and Anstey 2020; Tucker-Drob 2019 for reviews and meta-analysis). When observing changes longitudinally, researchers find that education only influences the level of cognition (or intercept) but not the change in (or slope of) cognition (see meta-analysis from Opdebeeck, Martyr, and Clare 2016; Seblova, Berggren, and Lövdén 2020). Researchers theorize the mechanism connecting education and cognition is through cognitive reserve, or the adaptability—capacity, efficiency, and flexibility—of the brain to stave off ADRDs and cognitive decline in old age by creating and maintaining denser neurological pathways in the brain (Stern 2002, 2012; Stern et al. 2020). Education is often used as a proxy for cognitive reserve because it is thought to either improve or reflect the experiences that contribute to cognitive reserve (Stern et al. 2020). Cognitive reserve is thus inextricably linked to education and has the potential to elucidate racial and gender differences in cognition and ADRDs.

Cognitive reserve provides a mechanism for how education translates into cognition: education provides the means for building cognitive reserve, while differential educational environments shapes who has access to cognitive reserve-building experiences (Berkman and Glymour 2006; Glymour and Manly 2008; Mungas et al. 2018; Stern et al. 2020). In the United States, past de jure and current de facto discrimination and segregation highly structured (and structure) educational attainment for people of color and women (Berkman and Glymour 2006; Glymour and Manly 2008). Especially for those who were school-aged before the Civil Rights Act of 1964, race and gender—and their intersection—dictate the level and quality of schooling one receives. In fact, evidence finds that Jim Crow era school segregation (Peterson et al. 2021)

and education quality (Liu et al. 2022; Sisco et al. 2015) explains a large portion of the race and gender differences in cognition, while others find that increased high school graduation rates from 1900 to 1950 lowered rates of dementia for both the Black and White population (Hayward et al. 2021). These findings suggest that access to education and education quality influence later life cognition through cognitive reserve and its differential distribution by race and gender.

Cognition and Geography in Old Age: A Consideration of the Southern United States

Geography proves to be an important factor in predicting health in the United States. There are clear associations between state of residence and health outcomes, particularly for those in the South (Montez and Farina 2021). Mortality, disability, and cardiovascular disease are significantly higher in the southern states compared to northern, western, and midwestern states (Fletcher et al. 2023; Kemp, Grumbach, and Montez 2022; Montez, Hayward, and Wolf 2017; Montez, Hayward, and Zajacova 2019). In fact, the contiguous states in the southeast United States have been dubbed the "Stroke Belt" because of the exceedingly elevated rates of and mortality due to cardiovascular disease (Howard and Howard 2020). Researchers find a similar geographic association with cognitive impairment in the South, such that residents in the South have a higher risk of dementia, cognitive impairment, and ADRDs compared to people who live outside the South (Ailshire, Walsemann, and Fisk 2022; Topping, Kim, and Fletcher 2021; Zacher, Brady, and Short 2023). Researchers have dubbed these contiguous states in the southeast U.S. as the "Dementia Belt" (Gilsanz et al. 2017; Zacher et al. 2023). Scholars have suggested that why southern states have high mortality and morbidity due to ADRDs is in part due to state-level policies. In the United States, routine legislation that most effects health and cognition are determined at the state level, including education reforms, Medicaid and Medicare

expansion, distribution of public assistance, civil investment, and economic policies like determining the minimum wage (Kemp et al. 2022; Montez et al. 2019). In addition to the power that states have to affect their residents' health and cognition, policies vary widely from state to state, and this variation translates into vast disparities in life expectancy and disability, including rates of ADRDs (Montez and Farina 2021; Montez et al. 2019; Montez, Hayward, and Zajacova 2021). Geography thus has a substantial influence on cognitive health through exposures to either generous or austere social, economic, and health policy.

In considering intersectional race-gender groups from the SDOH framework, exposures to state-level policies are influenced by historical and structural policies that differentially effect marginalized groups. More than half of the U.S. Black population lives in the South (Tamir 2021), resulting from historical enslavement and exploitation, de jure and de facto disenfranchisement, and migration patterns (Tolnay 2003). While differences in dementia incidence and cognitive impairment may at first appear to simply be a matter of population composition, the racial patterning of disadvantage is in fact more severe within the South. Researchers have found greater disparities in cognition and diagnosis of ADRDs between southern-residing Blacks and Whites compared to their non-southern counterparts (Liu et al. 2015; Zacher et al. 2023). The implication of these findings is that the combination of being Black and living in the South is particularly detrimental to cognitive health, where as White populations see no geographic difference in the incidence of dementia (Zacher et al. 2023). Ultimately, social-structural forces are at play in the influence that geography has on the racial patterning of cognitive decline.

HYPOTHESES: UTILIZING INTERSECTIONAL THEORY AND THE SOCIAL DETERMINANTS OF HEALTH TO PREDICT COGNITION TRAJECTORIES

The influence that allostatic load, education, and southern residence have on the unbiased age trajectories viewed through an intersectional lens offers a framework to understand how previous evaluations of cognition trajectories may be misspecified. Popularized by Kimberlé Crenshaw, intersectionality theory argues that multiple social statuses must be considered as existing simultaneously to acknowledge the unique intersecting, multiplicative systems of oppression operating on specific groups (Crenshaw 1989, 2017). While acknowledging Crenshaw's original impetus was to highlight systems of oppression and particularly the erasure of Black women's experiences, intersectionality theory widened the scope of how social scientists view power and inequality, especially in the quantitative social sciences (Bauer 2014; Carbado and Roithmayr 2014; Cho et al. 2013). Exposure to systems of oppression simultaneously prevent marginalized groups from developing greater cognitive reserve, while also bolstering access and opportunities for advantaged groups (Adkins-Jackson et al. 2023; Glymour and Manly 2008). This may be in the form of educational opportunities in early life (see Lenehan et al. 2015; Maccora, Peters, and Anstey 2020; Seblova, Berggren, and Lövdén 2020 for reviews) or cognitively stimulating work and hobbies in adulthood (see Fisher et al. 2017 for review). The components of interest in the following analyses are the latent intercept factor, latent slope factor, and the association between the intercept and slope factor with allostatic load, education, and southern residence. The comparison will be with the baseline model, or the latent growth model of cognition that has been adjusted for measurement non-invariance. The full model will control for allostatic load, education, and years lived in the South. The crux of these analyses is to evaluate the intercept, slope, and their changes for the baseline

and full model, and to compare the estimates from the two models to gain insight into how these covariates inform the trajectories for Black women, Black men, White women, and White men.

Allostatic Load

Hypothesis 1a: *Intercepts*

I hypothesize that when allostatic load (AL) is controlled for in the full model, the intercepts for Black women and Black men will increase compared to the baseline model, since controlling for AL removes the dampening effects AL has on cognition levels. In particular, I anticipate that, evidenced from work regarding the Superwoman Schema, that the intercept for Black women will increase more compared to Black men because of the multiple, intersecting forms of oppression Black women experience. I hypothesize this is a result of exposure to discrimination and marginalization, and subsequent cardiometabolic dysregulation, leading to greater prevalence of pathological levels of allostatic load. In contrast, I anticipate White women's and men's intercepts will remain unchanged between the baseline and full models. The alternative hypothesis is that there are no differences between each of the groups with regard to their allostatic load.

Hypothesis 1b: Slopes

Evidence for the relationship between AL and changes over time is not clear—there are few studies that investigate the influence of AL on cognition longitudinally. I suspect that the slope will remain unchanged for all groups, only influencing the initial level of cognition.

Education

Hypothesis 2a: *Intercepts*

I hypothesize the association between education and cognition will be positive across all groups and explain a large portion of the level of cognition across groups. This suggests the intercepts across all four groups will decrease in the full model compared to the baseline model. I expect that Black men and Black women will have smaller changes in their intercepts compared to their White counterparts due to structural forces that affected the quality and access of education earlier in their life course. The alternative is that the change between each of the groups is equal.

Hypothesis 2b: Slopes

Following extant literature, I anticipate there will be no difference in the slopes between the baseline model and the model controlling for education. Numerous studies have found that change in cognition is not a function of education (Opdebeeck et al. 2016; Seblova et al. 2020a). Alternatively, I hypothesize that the slopes may decrease for Black women utilizing evidence from the Double Jeopardy theory, where the double and unique challenge of living as Black women in a White patriarchy equates to worse, declining cognition when the buffer of education is removed.

Southern Residence

Hypothesis 3a: *Intercepts*

I hypothesize southern residence in old age will have the greatest impact on level of cognition for Black men and Black women due to the detrimental influence that living in the South has on the Black versus White population. This will translate to a negative relationship between the intercept and the number of years lived in the South, with elevated intercepts in the full model

compared to the baseline model. By comparison, I anticipate the intercepts for White women and White men will remain the same from the baseline to the full model, and will have statistically insignificant relationships between the intercept and southern residence. Alternatively, there may be no differences between living in the South across groups.

Hypothesis 3b: Slopes

Southern residence has the potential to augment the rate and severity of decline over age. Again, given the distinct health disparities between White and Black people in the South, I hypothesize that Black men and Black women will have larger negative slopes in the model controlling for southern residence compared to the baseline model due to sustained exposure. I anticipate the slopes of White men and White women will remain unchanged. Alternatively, Black women and men may have increased slopes as a result of removing the influence that the detrimental influence of Southern residence has specifically on the cognition of Black women and men.

METHODS

Data

The Health and Retirement Study (HRS) is a nationally representative longitudinal study of Americans over age 50. The HRS is an ideal data set for these analyses given its focus on aging populations and rich, detailed data collection. Respondents are surveyed once every two years and a new cohort is introduced every six years. The study draws its respondents using a probability sample of households in which at least one member is over 50 years old and non-institutionalized. The survey began in 1992, and recently released its 15th wave of data for 2020. The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740)

and is conducted by the University of Michigan. The present study uses data from 1996 (Wave 3—from which point cognition questions are asked consistently) to 2018 (Wave 14), the most recent wave for which sampling weights have been calculated (Health and Retirement Study 2022; RAND Corporation 2022). The sampling weights, stratum, and clustering variables account for the HRS's sampling structure, which oversamples Black populations, Hispanics/Latines, and Florida residents (Health and Retirement Study 2022; RAND Corporation 2022). These weights are used in the following the analyses. These data are publicly available online, in a harmonized, cleaned, and coded user-friendly dataset produced by RAND corporation (RAND Corporation 2022).

Dependent Variable

Cognition in the HRS is assessed using the HRS-Telephone Interview for Cognitive Status (HRS-TICS). The twelve tests in the HRS-TICS measures two umbrella concepts: episodic memory and mental status. The TICS uses assessments to measure episodic memory: an immediate word recall (IMRC) and delayed word recall (DLRC) test. Respondents are given a list of ten words from four possible lists and are asked to immediately recall as many words from the list to the best of their ability (IMRC) and to complete the same task five minutes later (DLRC) (McArdle et al. 2007). Scores range from 0 and 10 for both tests.

The remaining cognitive tests measure mental status, and include: serial 7s test (count backward from 100 by 7 for 5 trials; scored 0-5); counting backwards from 20 for ten continuous digits in two trials if needed (scored 0-2); identifying a cactus and a pair of scissors after listening to a description (scored 0 or 1 for each; scored 0-2); correctly naming the current president and vice president (scored 0 or 1 for each; scored 0-2); and stating the day's date (day,

month, year, day of the week; scored 0-4). Frequently, scholars sum together the memory and mental status scores to create a composite cognition score with possible values from 0 to 35 (McArdle et al. 2007; McCammon et al. 2022; Ofstedal et al. 2005).

For the purposes of this study, I recode the cactus, scissors, president, and vice president dummy variables into a single four-point ordinal categorical variable (referred to as "Names"). The pursue the same strategy for the date dummy variables, referred to as "Dates". The recoding is necessary because these two sets of dummy variables are highly correlated and fit the data better as two ordered categorical variables as opposed to eight dummy variables. Lastly, the counting backwards from 20 variable is recoded to a dummy variable, where respondents who successfully completed the test in one or two trials are assigned "1" (named "CNTB"). I recoded this because very few people were unable to count backwards from 20 on their first try, and thus would not contribute to the measurement model of cognition. More substantively, I wanted to differentiate between who was able versus unable to complete the task, and not necessarily to what degree respondents were able to complete the task. Notably, the Names and Dates variables are recoded as described in other foundational works on measurement properties of the HRS's cognitive function factor (McArdle et al. 2007; Ofstedal et al. 2005).

Predictor Variables

The intersectional identities of particular interest in this chapter are respondents' race, ethnicity, and sex (referred to as gender hereafter). Race is measured as Black, White, or other in the HRS. I exclude in this study respondents who identified as other, leaving only those who self-identified as Black or White. I made this choice because the Black and White categories make up the majority of racial identities in the HRS and, more importantly, the heterogenous

"other" category does not have racial distinctions in the Public Use data; if I were to include it in my analysis, I would not know which racial(ized) category respondents belong to.

Ethnicity is measured as self-identifying as Hispanic or non-Hispanic. I excluded those who identified as Hispanic because of the heterogeneity in Hispanic origin, which is only crudely categorized as "Mexican" and "other" in the Public Use data. I made this decision on the basis that people who identify as non-Hispanic White and non-Hispanic Black are less likely to speak English as a second language. If I were to introduce Hispanic/Latine respondents, it is likely that the source of measurement non-invariance would be due to language barriers, as opposed to focusing on bias due to the hegemony of White men. A crosstab of the race and ethnicity variables resulted in the two race/ethnicity groups of interest in this study: non-Hispanic White (referred to as White hereafter) and non-Hispanic Black (Black hereafter).

Gender was measured as a binary, either male or female. I acknowledge that conceptualizing gender beyond a simple binary is a worthwhile investigation in terms of measurement invariance and is well established in gender studies as a social status that brings with it its own systems of oppression. However, besides the fact that the HRS only provides "female" and "male" answer options, my theoretical motivation for this paper relies primarily on the health impacts of (White) male hegemony. Future studies ought to include measurement bias related to non-binary or transgender respondents. The focal intersecting social statuses in this chapter are race/ethnicity (Black and White) and gender, which combine to create four sub-groups: Black men, Black women, White women, and White men. These race-gender intersectional groups represent the social statuses that either (in the case of White men) benefit from or (in the case of the remaining three groups) are harmed by the primacy of White patriarchy and its centrality in cognition assessments.

Age is the fundamental axis on which cognition changes and, thus, is the central independent variable with regard to changes in cognition. While this presents a missing data issue—there are far fewer people missing across 12 waves of data compared to 50 age categories—Mplus, by default, uses a full information maximum likelihood (FIML) to estimate the models (Muthén and Muthén 2017). The benefit of FIML is that it produces unbiased estimates and standard errors even with missing data (assuming data is missing at random [MAR]) by estimating a likelihood function for each individual using all available data (Enders and Bandalos 2001). The HRS conducts a full cognition assessment at respondents' initial survey, then again when the respondent is at least 65, and at every wave thereafter (Ofstedal et al. 2005). Between their initial wave to age 65, the respondents are given a partial battery, which only includes immediate (IMRC) and delayed word recall (DLRC), counting back from 20 (CNTB), and counting back from 100 by 7s (Serial 7s) (Ofstedal et al. 2005). To further minimize the aforementioned missing data issue, the sample is restricted to people aged 65 to 80 to avoid the empty cells for the full battery between age 50 to 65. Finally, because respondents were surveyed once every two years, the data were collapsed into two-year age categories starting with age 65-66 up to age 79-80. However, because the HRS did not necessarily interview respondents within the wave year, respondents often were grouped twice in one category because the interview date was less than two years apart from the previous interview. To ensure that respondents had only one observation per two-year age interval, I recoded age to reflect the age that respondents were in the wave year, not the interview year. This resulted in every respondent being in one two-year age category to maximize the sample size and minimize missing data; however, this led to ~25% of the sample to be in an age category one year higher than their actual age. This is the most logical data format since the shift up in age is presumably random

and given that the alternative would require separate age categories which would leave one-year age gaps across a majority of respondents.

In addition to age, I control for contextual factors that influence cognitive decline over age: allostatic load, years of education, and number of years lived in the US South. The HRS began collecting dried blood spots, physical measures, and blood pressure on a random sample of half the respondents in 2006 and the other half in 2008 (Crimmins et al. 2020). The two samples had dried blood spots, physical measures, and blood pressure taken every other wave or every four years (i.e., in 2006, 2010, and 2014 or in 2008, 2012, and 2016). The dried blood spots were assayed for five biomarkers: total cholesterol, high density lipoprotein (HDL) cholesterol, glycosylated hemoglobin (HbA1c), c-reactive protein (CRP), and cystatin C (Crimmins et al. 2020).

Total cholesterol and HDL cholesterol indicate the amount of total and "good" cholesterol; implicitly, the difference between total cholesterol and HDL cholesterol is low density lipoprotein (LDL) cholesterol, or so-called "bad" cholesterol because of its association with heart disease and stroke (Center for Disease Control and Prevention 2023). HbA1c reflects the average free-floating glucose in the blood over the previous three months; at higher levels, serum glucose indicates glycemic dysregulation and an elevated risk of diabetes (World Health Organization 2011). CRP is an acute-phase serum protein that the liver produces in response to inflammation (e.g., sickness, injury), as well as stress (Steptoe, Hamer, and Chida 2007). Cystatin C is a protein produced by virtually all cells and, at high levels, indicates poor kidney function, which increase the risk of cardiovascular disease (Inker and Levey 2018). Lastly, elevated pulse and blood pressure measurements indicate cardiovascular issues, while waist

circumference measures abdominal adiposity, which is associated with cardiometabolic syndromes (Crimmins et al. 2008).

Together, these nine biomarkers reflect the accumulation of physiological dysregulation and the associated so-called "wear and tear" on the body's biological systems (B. S. McEwen 2016; Schmitz et al. 2018). According to a systematic and meta-analytic review by D'Amico and colleagues (2020), the most common way to operationalize these variables in the HRS is to create allostatic load index using a quartile count. Higher levels of each biomarker indicates higher allostatic load, except HDL, which was reverse coded. Respondents above the 75th percentile of each biomarker were coded as one and all those below the threshold were coded zero. Quartiles were calculated across the full sample to ensure the representativeness of the U.S. population (Oi and Haas 2019). The dummy variables were then summed, creating an index of zero to nine, where higher counts indicate a higher allostatic load. I then calculated the mean value of the index to create a time-invariant mean allostatic load variable. The measure of allostatic load I utilize represents the respondent-specific level of physiological dysregulation, which at higher levels indicates the potential for worse cognition.

Years of education are measured as a continuous variable from 0 to 17, which will allow for easier interpretability of model results. Here, education is a proxy for, or at least representative of the processes that increase, cognitive reserve (Salthouse 2009). More years of education will indicate greater cognitive reserve and thus higher levels of cognition. Lastly, I used the HRS publicly available data on which Census region respondents live in at each wave to measure the number of years that each respondent has lived in the southern United States. The southern Census region includes Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi,

Tennessee Arkansas, Louisiana, Oklahoma, and Texas. The measure is of the cumulative exposure to the south's policy environments, which lack the social, economic, and health policies that stave off ADRDs and improve cognition outcomes (Gilsanz et al. 2017; Montez and Farina 2021; Zacher et al. 2023). By modeling residence in the south as an exposure, as opposed to a dummy variable, I am able to quantify and model how living in the southern context influences the cognition trajectories across the four intersectional race-gender groups.

Analytic Plan

I conducted a multi-group second order (or multiple indicator) latent growth model (SGM) within a structural equation modeling (SEM) framework to test for measurement invariance of cognition by intersectional race-gender groups over age. Substantive measurement invariance analyses were conducted using Mplus version 8.8 (Muthén and Muthén 2017). The SGM differs from the better-known first order (or single indicator) latent growth curve model (LGCM) in that there are two parts: the measurement model and the latent growth model. The advantage of this two-step process is that it allows researchers to test for measurement invariance of the latent factor over repeated measures and over groups (Chan 1998; Geiser et al. 2013; McArdle 1988). Building off of chapter one, I estimate separate measurement models for each two-year age group for eight age groups (65-66, 67-68, 69-70...79-80), focusing only on the scalar and partial scalar models from my previous chapter. The second step utilizes the latent cognition factors at each time point from the first step and models them in a LGCM over eight two-year age groups. The resulting LGCM estimates a latent intercept factor and latent slope factor for each race-gender group. These latent intercepts and slopes represent the growth trajectories, or estimated underlying growth process that leads to differential trajectories in the

latent factor (Kline 2015; McArdle and Epstein 1987; Meredith and Tisak 1990). In this chapter, the intercept represents the average latent cognition factor for each race-gender group at age 65 to 66. The latent slope factor measures the average change in the latent cognition factor over age, measured here in two-year intervals. The LGCMs in this chapter are linear models with continuous outcomes.

Before running an SGM, the measurement model must first be identified—as is, there are more parameters than there are unique values to predict them (Kline 2015). In order to identify the model, the cognition factor requires a scale since latent factors lack an inherent metric (Sass 2011). According to experts, the scaling variable should not be random, but based on which item in the measurement model is "most invariant" (Cheung and Rensvold 1999; Sass 2011), which in this case is the IMRC item. The factor loadings for the IMRC item are set to one across all race-gender groups. The remaining factor loadings for DLRC, CNTB, Ser7, Names, and Dates are time-invariant and set to that of the reference group (White males) across the intersectional race-gender groups. In addition, the intercepts for the IMRC item are set to zero at each time point and across all race-gender groups. The two models of interest—the scalar and partial scalar models—differ in how their remaining intercepts are modeled. In the scalar model, the remaining items (DLRC, CNTB, Ser7, Names, and Dates) time-invariant and are set to that of the White male group for all of the race-gender groups. In the partial scalar model—the specification of which I determined in Chapter 1—the intercepts remain time-invariant, while the DLRC and CNTB intercepts are set to that of the White male group and the Ser7, Names, and Dates intercepts are freely estimated across the intersectional race-gender groups. The time-invariance of the scalar and partial scalar model is a substantive decision given current evidence finds that the cognition factor is invariant over age and time (Avila et al. 2020; Barnes et al. 2015;

Blankson and McArdle 2015), meaning the relationship between item factor loadings and intercepts with the latent cognition factor do not change over age or time.

As previously noted, the focus of this chapter is on the scalar and partial scalar SGMs. An important facet of SGMs is the model fit, which conveys how well the measurement model and LGCM fit the data across intersectional race-gender group (Ferrer et al. 2008). Similar to the multi-group confirmatory factor analysis of the previous chapter, I determine whether model fit improves between the scalar and partial scalar SGMs models. I utilize three goodness-of-fit indices: the χ^2, the root mean square error of approximation (RMSEA), and the Comparative Fit Index (CFI). The χ^2 is a general fit index, but is heavily influenced by sample size and may show statistically significant differences in model fit when in fact the model fit differences are negligible (Chen 2007; Cheung and Rensvold 2002). As Cheung and Rensvold (2002) recommend, I focus on the RMSEA and CFI as indicators of model fit because they are superior indices for detecting non-invariance in complex models with large sample sizes. Substantively, the RMSEA evaluates the degree to which the estimated model differs from that of a theoretical fully saturated model, where all variance and covariance is explained, thus numbers closer to 0 indicate better fit. The CFI, on the other hand, measures the discrepancy between the estimated model and a theoretical baseline model where none of the variance and covariance is explained, thus a number closer to 1 indicates a better fit (Lai and Green 2016). To employ these model fit statistics, I utilize common cutoffs that previous experts have recommended, specifically applying threshold values at or below 0.05 for the RMSEA and values at or above 0.95 for the CFI to indicate a well-fitted model (Svetina et al. 2020; Wilson et al. 2023). To gauge whether model fit improves between the more restrictive scalar 2LGCM and less restrictive partial scalar 2LGCM, I utilize the ΔRMSEA and ΔCFI, which are simply the differences between the indices

from two models. To determine if the model fit changed significantly, values of ΔCFI less than -0.005 or ΔRMSEA values greater than 0.010 indicate poorer fit and evidence that invariance at that level does not hold (Chen 2007; Cheung and Rensvold 2002; Svetina et al. 2020).

In order to control for the three contextual factors of interest, I will additionally include a regression of the latent intercept and slope factors on allostatic load, education, and years lived in the South. The output includes the coefficients from these regressions, which indicate the average change in the latent intercept and slope factor with every one unit increase in allostatic load, education, and years lived in the south. The latent intercept and slope factors in regressions controlling for these three covariates reflect the average latent cognition factor intercept and slope when the variation due to allostatic load, education, and years lived in the South are removed. The results will convey in which direction and to what degree these three covariates influence respondents starting cognition (at age 65-66) and average change in cognition over each two-year interval. The resulting latent intercept and slope factors from the baseline and full models reflect both the estimated trajectories when bias due to measurement non-invariance is removed and, in the full model, when the addition of three important individual- and contextual-level variables that influence cognition are included in the model.

Current Study Sample

The starting sample in the HRS for Waves 3 (collected in 1996) to Wave 14 (collected in 2018) is N = 232,326 individual observations among 39,958 respondents. In order to be included in the analytic sample, respondents had to have non-missing values for race, ethnicity, gender, sampling weights, and had to be aged 65 to 80. Respondents had to respond to the survey themselves and not by proxy. Additionally, respondents and had to either be in the 2006 or 2008

waves to submit a dried blood sample and have non-missing values for the biomarker and biometric data. Lastly, respondents had to have a non-missing value for years of education and the census division. The final sample size is N = 9,698 respondents, with 500 Black men, 920 Black women, 3,428 White men, and 4,850 White women.

Summary sample statistics include unweighted means and standard deviations of the allostatic load index, years of education, and years lived int the South (Table 1) for the full sample and across race-gender intersectional group. All summary and descriptive statistics were conducted using Stata 17 (Stata Corp 2021). The average allostatic load index, which ranges from 0 to 7, was around 2.0 across the full sample and separately by intersectional race-gender group. This reflects that, on average, everyone in the sample fell in the top 75th percentile in two of the nine biometric variables. For the full sample, the average years of education was approximately 13 years. By race-gender group, average years of education was lowest for Black men (11.1 years), followed by Black women (11.7 years), White women (12.9 years), and White men (13.3 years). Lastly, the average years that people lived in the South, starting at age 65, was 4.1 for the full sample. There is a clear racial distinction in southern residence, with Black men and women living in the South for an average of 6 years compared to White men and women, who average around 4 years living in the region. The average years of education and southern residence provides context for the following analyses, where there is a clear gradient across race-gender groups with regard to education and years lived in the South. The differences in education and southern residence may partly explain part of the variation in race-gender cognition trajectories, namely through differential education opportunity structures and exposures to austere state policy environments.

RESULTS

Model Results of Second-Order Latent Growth Models Controlling for Contextual Variables

The following results focus on the baseline and full model with all three covariates included (Table 1). Models controlling for each variable separately and combination are in the Supplementary Table 1. The baseline model results (Table 1) include the latent intercept and slope over the four intersectional race-gender groups. The baseline model does not include any covariates and adjusts the measurement portion of the growth model to be invariant across race-gender groups. The latent intercepts, which reflect the average latent cognition score when respondents are 65-66 years old, reflect the race-gender trends outlined in my previous chapter: White women have the highest intercept (6.37), followed by White men (5.82), Black women (5.54), and Black men (5.01). All of the intercepts are statistically significant at the $p = 0.001$ level. Additionally, the 95% confidence intervals of the intercepts do not overlap between groups, meaning the intercepts are statistically distinct from one another. Previous research that utilizes growth modeling (outside the SEM framework) find cognition trajectory gaps are solely drawn down racial lines, and see no differentiation by gender (e.g., Yang et al. 2023). This may be indicative of the SGM's ability to create equivalent measures of cognition for Black and White women to that of Black and White men over age. The latent slopes, which present the average linear change in cognition over each two-year period, are virtually the same across intersectional race-gender group, between -0.14 to -0.12. Because these estimates are close together, the 95% confidence intervals overlap with one another, and thus are not statistically distinct from one another. These slopes do mimic linear declines in cognition from middle to late adulthood (e.g., Salthouse 2009, 2019; Yang et al. 2023), suggesting that the measurement invariance analysis does not alter the rate of change in cognition. These results differ from what

researchers have found when looking at age trajectories over race-gender groups, where there is a clear distinction between each group, with White women and Black women outperforming their male counterparts.

The full model results on the right side of Table 1 present the findings for the multiple-indicator latent growth model when I control for mean allostatic load, years of education, and number of years lived in the south. The resulting intercept and slope in the full model are the estimates of the intercept and slope devoid of variation due to allostatic load, years of education, and living in the South. The intercepts decrease significantly in the full model compared to the baseline model, where White women have the highest intercept (4.10), followed by White men (3.45), Black men (3.11), and Black women (3.02). As compared to the baseline model, the 95% confidence intervals for the intercepts overlap between Black men and Black women; the confidence intervals do not overlap between White men and White women, suggesting the three covariates do not account for the difference in intercepts between White men and White women. As for the slopes, there is a clear increase in estimates in the full model compared to the baseline model for White men (from -0.14 to -0.08), White women (from -0.13 to -0.07) and Black women (from -0.14 to -0.08). Black men, on the other hand, have a slope of -0.14 in both the baseline and full models. The slopes estimates are all statistically significant at the p=0.05 levels, except for Black women, whose slope is not statistically different from zero. As evidenced in Figure 1, when I include the three covariates in the model, the resulting trajectories have minimal differences between Black women, Black men, and White men. It appears that with the addition of adjusting for measurement non-invariance, controlling for allostatic load, education, and southern residence nearly eliminates the race-gender differences. What led to the compression of

these trajectories is evident in the regression coefficients of the intercept and slope on allostatic load, education, and southern residence.

The regression coefficient of the latent intercept factor on allostatic load was statistically significant for White men and women, where a one unit increase of average allostatic load was associated with a -0.06 and -0.04 unit decrease in cognition, respectively. The relationship between average allostatic load and the latent intercept for Black men and women was not statistically significant. This runs in contrast to the Hypothesis 1a, where AL has a negative relationship only for White women and men and not for Black women or Black men. In controlling for AL, the intercept factor slightly increases, removing the negative association between AL and cognition for White women and men. However, there is evidence for Hypothesis 1b, where the regression of the slope on AL was not statistically significant across all groups.

Education had a statistically significant association with the latent intercept factor across all four groups, which explains the drastic decrease in intercepts between the baseline and full models. For White men and White women, a one-year increase in education is associated with approximately 0.17 unit increase in cognition. For Black men and Black women, a one-year increase in education is associated with a 0.22 and 0.25 unit increase in the intercept, respectively. These results run contrary to Hypothesis 2a. White women and men have a smaller association between the intercept and education compared to Black women and men, suggesting there is a slightly greater "conversion" of educational attainment into higher cognition for Black women and men. This runs contrary to research on the association between education and cognition, which usually finds the opposite—Whites have a larger association between education and cognition compared to Black men and women because they are able to translate education

into higher cognition through accessing the benefits that resourced social networks, neighborhoods, and health care can provide (Avila et al. 2021; Yang et al. 2023). The association between the latent slope factor on years of education was statistically significant for Black women, where a one-year increase in education was associated with a -0.01 unit decrease in the latent slope factor. This finding is seemingly counterintuitive, though this may reflect a faster decline in cognition given the higher intercept that higher education would confer. The remaining groups had no association between the slope factor and education. The findings regarding the relationship between the intercept and slope with education replicate what other scholars have found in the HRS and other datasets of aging Americans (see Lövdén et al. 2020 and Seblova, Berggren, and Lövdén 2020 for meta-analyses and reviews) for Black men, White women, and White men. Education is associated with intercept, or level, of cognition and not the slope, or change over time. This which falls in line with the prediction of Hypothesis 2b, but does not hold for Black women.

The regression of the average years lived in the South were not statistically significant across all groups in the full model. However, in Supplemental Table 1, the model that only includes the southern residence variable finds a negative significant relationship for Black women only. This suggests that living in the south for Black women is detrimental to the level cognition, but not change in cognition. The fact that the association disappears in the full model (Table 1) suggests that years of education moderates the relationship between southern residence and cognition—education may be the modifiable element that makes exposure to the policy environments in the South inconsequential for cognition levels for Black women in this sample.

A comparison of the baseline and full model trajectories are plotted in Figure 1. The baseline model trajectories have clear, wide gaps in the level of cognition, but little difference in

the slope, or change, in cognition over time. The full model—which accounts for measurement invariance and contextual factors—nearly eliminates the disparities between Black men, Black women, and White men. Controlling for allostatic load and education in addition to adjusting for measurement non-invariance compresses the trajectories together. Education appears to account for most of the difference in the intercept, which decreases drastically between the baseline and full models. The slopes for Black women, White women, and White men shift towards zero by more than half, however Black men's slope remains the same between the baseline and full models. The reason for this change in slope appears to be due to the large shift in the intercepts across all groups moving down by several units. While the disparities are mostly accounted for among Black women, Black men, and White men, the gap between White women's cognition trajectory and the three other trajectories remains.

Fit Indices for Baseline and Full Model

The model fit indices convey how well the data fit the structural equation I modeled in the second-order latent growth model across the intersectional race-gender groups (see Chapter 2). An improved change in the fit statistics from the baseline to the full model will suggest that the inclusion of covariates at the individual, institutional, and structural level reflects a more accurate portrayal of the data. The fit of the baseline model (Table 2) reflects how well the data fit when the cognition factor loadings and intercepts are time-invariant; the cognition factor loadings are set to the reference group (White men); the intercepts for DLRC and CNTB are set to the reference group; and the intercepts for Ser7, Names, and Dates are freely estimated. The model fit indices are mixed with regard to how well the growth model fits the data. The RMSEA is 0.019, which suggests a well-fitted model, with the fit statistic falling under the threshold of 0.05. However, the baseline model's CFI is 0.897, which is well below the threshold for a well-

fitting model (0.95) but very close to an adequate fit (0.90). In the full model, the latent growth factor intercept and slope are regressed on variables for allostatic load, years of education, and total years lived in the south. Similar to the baseline model, the fit indices point to a semi-well-fitted model, with an RMSEA of 0.017 and a CFI of 0.852. The change in fit between the baseline and full models suggests that there is mixed evidence for a change in fit, with a ΔRMSEA of 0.002 and a ΔCFI of 0.045. The ΔRMSEA suggests that the model fit improves slightly and ΔCFI suggest that the model fit deteriorates significantly. However, experts suggest that in confirmatory factor analysis, the RMSEA should be considered with greater weight than the CFI because the former is the comparison between a fully saturated model (i.e., all the covariance explained), which is presumably the goal of SEM (Lai and Green 2016; Rigdon 1996). In addition, the cutoff points for goodness-of-fit indices are arbitrary (Lai and Green 2016; Svetina et al. 2020); given that the results largely mimic previous findings and logically follow the results of my previous chapters adds to my confidence that the model fit is acceptable and slightly improves with the addition of the three covariates.

DISCUSSION

Previous research has established that cognition in aging Americans is discrepant across race and gender, where Black people's cognition fairs worse than that of White's, while women bear the brunt of Alzheimer's disease in old age compared to men. In trajectories over age, researchers have attempted to explain the gaps between race and gender groups by controlling for relevant social factors, like education, SES, stress, and geography, but often still find a race- and gender-based gaps. However, previous studies on race- and gender-based disparities in cognition trajectories over age have neglected the potential for bias in the cognition assessment

themselves. A potential answer lies in utilizing second-order latent growth curve models (SGM) within a structural equation modeling (SEM) framework, where analysts can assess if and to what extent are cognition trajectories over age measured equivalently across social dimensions.

Most research on race and gender differences in cognition is executed under the assumption of measurement invariance. Measurement invariance is a foundational assertion that the latent variable of interest—here, cognition—is captured by the same set of tests and questionnaire items over social dimensions and time (Horn and McArdle 1992). However, it may be that some proportion of the differences in cognition are not real differences in cognitive abilities but an artifact of measurement non-invariance, where survey items and tests do not capture the same underlying construct across sub-groups (Horn and McArdle 1992). Extant research has not adjusted cognition trajectories for measurement non-invariance while simultaneously controlling for relevant covariates. The remaining disparities in cognition trajectories are possibly due to measurement non-invariance across race and gender group. The analyses of this chapter included allostatic load, education, and number of years lived in the South as social, contextual variables pertinent to race-gender cognition disparities in old age.

My analyses find that, in addition to adjusting for measurement non-invariance, allostatic load and education account for much of the difference in cognition trajectories across intersectional race-gender groups. Black women's, Black men's, and White men's cognition trajectories in old age are compressed together in the full model controlling for AL, education, and southern residence. The association between allostatic load and the intercept was not statistically significant for Black women or Black men, but had a negative association for White women and White men. This runs contrary to Hypothesis 1a, in which I hypothesized that Black women would have the largest association followed by Black men, and that there would be no

association for White women and men. The reason for the statistical relationship for the White sample may be because White women and men do not utilize adaptive coping mechanisms to the same extent that Black women and men do as a matter of survival in society organized by a White patriarchy (Perez et al. 2023), and thus have hypersensitive reactions to stressors. In fact, using a sample of Black men and women from the HRS, McDonough, Byrd, and Choi found that resilience resources—social support and social contact—buffered the effects of discrimination on cognitive decline (2023). There may be social resources or resilience that White men and women do not utilize and thus have elevated allostatic load. Additionally, Boen and colleagues found in their decomposition of the Black-White disparities in biological aging that structural, socioeconomic influences accounted for a majority of the disparity, while psychosocial (i.e., stress and allostatic load) factors accounted for a minimal proportion of the Black-White disparities in biological aging (2023).

Education appears to have a larger influence on the level of cognition compared to allostatic load. There is a clear positive, significant association between the intercept and education across all four groups. This partially supports Hypothesis 2a; however, the coefficients for the regression of the intercept on education for Black women and men were significantly higher than the coefficients for White women and men. These findings conflict with research that finds the opposite relationship between race, education, and cognition: White people are able to "convert" their education into better cognitive outcomes, while Black people experience diminishing returns to their education (Avila et al. 2021; Yang et al. 2023). However, the discordant results of this chapter may be due to adjusting the model for measurement non-invariance, which leads to differential changes to the relationship between education and cognition. The slope factor does change for Black women, White women, and White men,

shifting up towards zero despite the lack of association between the latent slope factor and education. This is perhaps due to the intercepts so drastically falling in the full model compared to the baseline model. Black men's cognition slope remained the same despite also seeing a drop in their intercept. Overall, by accounting for measurement non-invariance and the relevant covariates, I nearly eliminated the disparities in cognition trajectories. However, a gap remains for White women, who have an elevated trajectory compared to that of Black women, Black men, and White men. This mimics the findings from McDonough et al. who found that, when utilizing a different method of removing bias from cognition, White women had similarly persistent higher cognition compared to Black, Hispanic, and Asian men and women and White men (2022).

As with any analyses, this study comes with some limitations. For one, the allostatic load and southern residence are time-varying covariates, but I operationalized them as being time-invariant. Cognition may be more sensitive to immediate changes in stress levels and policy contexts, especially given the impact that allostatic load and health-related state policy has on the cognition of the aging population. Additionally, as with any study investigating aging populations, mortality selection possibly altered the sample, where people with very low cognition died, leaving a more cognitively robust sample. However, this is a perennial issue in aging research and may be circumvented by controlling for whether a respondent dies while in the study. Despite these limitations, these findings suggest that disparities in cognition trajectories in old age can potentially be accounted for through two means: incorporating measurement invariance analysis and including relevant covariates. Without viewing these trajectories by intersectional race-gender groups over age, these disparities would go unnoticed. Future analyses of cognition trajectories across socially defined groups must take into account

the potential for measurement bias in cognitive measurement tools when modeling changes over age.

TABLES AND FIGURES

Table 1. Unweighted descriptive statistics of the Health and Retirement Study 1996-2018, N=9,698 aged 65-80

		Allostatic load index	Years of education	Years lived in the south
Total	Mean	1.9	12.8	4.1
N=9,698	SD	1.2	2.7	4.9
100.0%	Min	0	0	0
	Max	7	17	12
Black men	Mean	2.2	11.1	5.7
n=500	SD	1.3	3.5	4.9
5.2%	Min	0	0	0
	Max	7	17	12
Black women	Mean	2.2	11.7	5.8
n=920	SD	1.3	3.0	5.0
9.5%	Min	0	0	0
	Max	6	17	12
White men	Mean	2.0	13.3	3.8
n=3,428	SD	1.2	2.8	4.8
34.3%	Min	0	0	0
	Max	7	17	12
White women	Mean	1.8	12.9	3.8
n = 4,850	SD	1.2	2.4	4.8
50.0%	Min	0	0	0
	Max	7	17	12

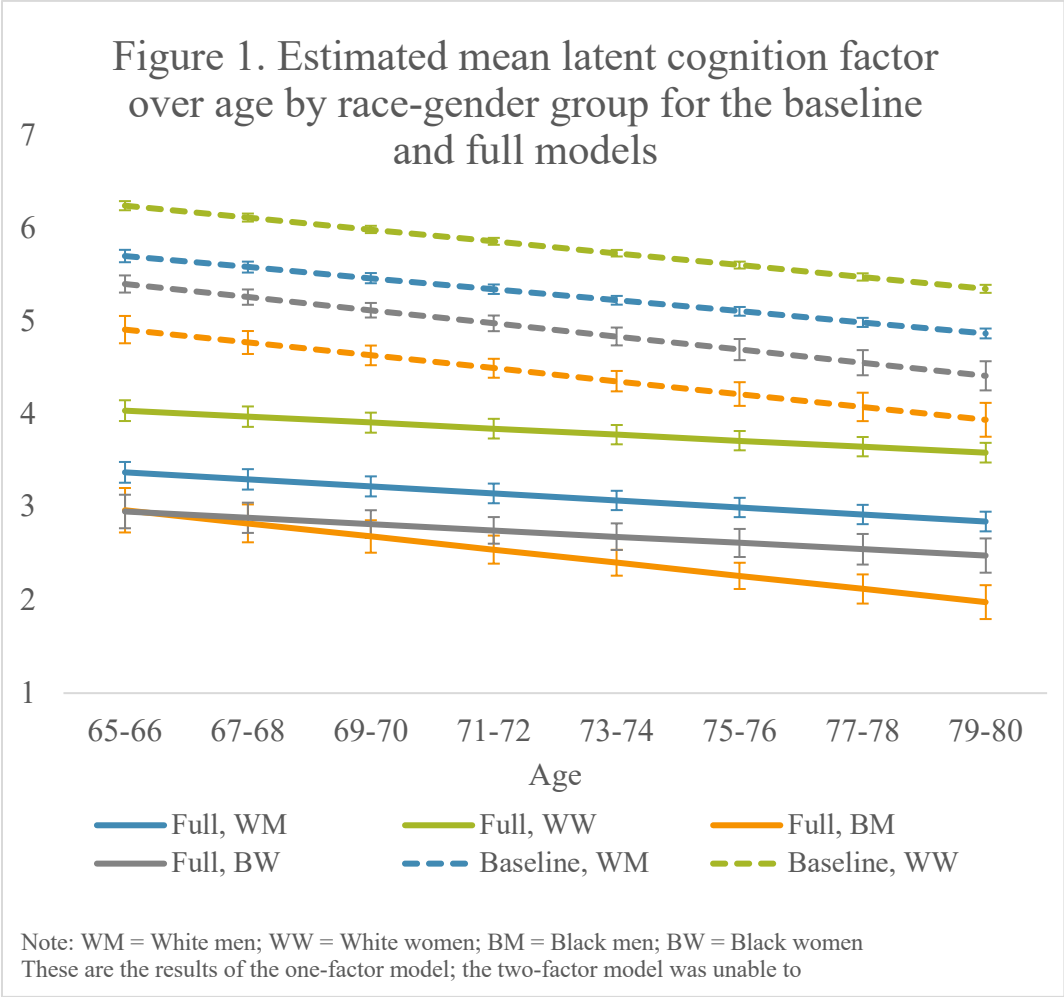
Table 2. Second-order latent growth model estimates for the baseline and full model estimates (& SEs) across race-gender group from the Health and Retirement Study 1996-2018, N=9,698 aged 65-80

	Baseline				Full model			
	WM	WW	BM	BW	WM	WW	BM	BW
Intercept	5.82***	6.37***	5.05***	5.54***	3.45***	4.10***	3.11***	3.02***
	(0.04)	(0.03)	(0.09)	(0.05)	(0.16)	(0.14)	(0.40)	(0.31)
Slope	-0.12***	-0.13***	-0.14***	-0.14***	-0.08**	-0.07**	-0.14*	-0.07
	(0.01)	(0.00)	(0.02)	(0.01)	(0.03)	(0.02)	(0.07)	(0.05)
Intercept on AL	—	—	—	—	-0.04†	-0.06***	0.00	-0.06
	—	—	—	—	(0.02)	(0.02)	(0.07)	(0.05)
Slope on AL	—	—	—	—	0.00	0.00	0.02	0.00
	—	—	—	—	(0.01)	(0.00)	(0.01)	(0.01)
Intercept on education	—	—	—	—	0.18***	0.17***	0.22***	0.25***
	—	—	—	—	(0.01)	(0.01)	(0.03)	(0.02)
Slope on education	—	—	—	—	0.00	0.00	0.00	-0.01†
	—	—	—	—	(0.00)	(0.00)	(0.00)	(0.00)
Intercept on living in south	—	—	—	—	0.00	-0.01	-0.01	-0.01
	—	—	—	—	(0.01)	(0.00)	(0.02)	(0.01)
Slope on living in south	—	—	—	—	0.00	0.00	0.00	0.00
	—	—	—	—	(0.00)	(0.00)	(0.00)	(0.00)

Note: WM = White men, WW = White women, BM = Black men, BW = Black women; ***p <0.001 **p<0.01 *p<0.05 †p<0.10; AL = average allostatic load (AL); education = continuous measure of years of education; living in the south = Average number of years living in the south since age 65

Table 3. Goodness-of-fit indices for the baseline and full models

	Baseline model	Full model
Chi-square	7795.53	8516.23
df	4252	4803
RMSEA	0.019	0.018
ΔRMSEA	—	0.001
CFI	0.897	0.873
ΔCFI	—	0.024



Note: WM = White men; WW = White women; BM = Black men; BW = Black women
These are the results of the one-factor model; the two-factor model was unable to

SUPPLEMENTAL TABLES AND FIGURES

Supplemental Table 1. Second-order latent growth model estimates for the baseline and full model estimates (& SEs) across race-gender group from the Health and Retirement Study 1996-2018, N=9,698 aged 65-80

	Allostatic load (AL) model				Education model				Live in South model			
	WM	WW	BM	BW	WM	WW	BM	BW	WM	WW	BM	BW
Int.	6.02***	6.56***	5.08***	5.75***	3.28***	3.92***	2.92***	2.92***	5.80***	6.39***	5.14***	5.68***
	(0.07)	(0.05)	(0.18)	(0.13)	(0.14)	(0.15)	(0.29)	(0.29)	(0.04)	(0.04)	(0.16)	(0.08)
Slope	-0.11***	-0.13***	-0.16***	-0.16***	-0.07***	-0.06***	-0.11*	-0.11	-0.12***	-0.13***	-0.12***	-0.16***
	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)	(0.03)	(0.02)
Int on AL	-0.12***	-0.13***	0.01	-0.13*								
	(0.02)	(0.02)	(0.08)	(0.05)								
Slope on AL	0.00	0.01	0.01	0.01								
	(0.00)	(0.00)	(0.01)	(0.01)								
Int on ed					0.18***	0.18***	0.22***	0.22***				
					(0.01)	(0.01)	(0.02)	(0.02)				
Slope on ed					0.00	0.00	0.00	0.01†				
					(0.00)	(0.00)	(0.00)	(0.00)				
Int on South									0.00	-0.01	-0.02	-0.04**
									(0.01)	(0.01)	(0.02)	(0.01)
Slope on South									0.00	0.00	0.00	0.00
									(0.00)	(0.00)	(0.00)	(0.00)

Note: Int. = Intercept; AL = Allostatic load; Ed = Years of education; South = Years lived in the South
WM = White men, WW = White women, BM = Black men, BW = Black women; ***p <0.001 **p<0.01 *p<0.05 †p<0.10; Model 1 controls for average allostatic load (AL); Model 2 controls for years of education; and model 3 controls for proportion of time spent in the south.

Supplemental Table 2. Second-order latent growth model estimates for the baseline and full model estimates (& SEs) across race-gender group from the Health and Retirement Study 1996-2018, N=9,698 aged 65-80

	Allostatic load (AL) + Live in South				Education + Live in South				Allostatic load (AL) + Education			
	WM	WW	BM	BW	WM	WW	BM	BW	WM	WW	BM	BW
Int.	6.02***	6.58***	5.20***	5.93***	3.27***	3.94***	3.10***	2.92***	3.45***	4.09***	2.96***	2.92***
	(0.07)	(0.06)	(0.23)	(0.14)	(0.14)	(0.15)	(0.35)	(0.28)	(0.16)	(0.13)	(0.32)	(0.28)
Slope	-0.12***	-0.13***	-0.15***	-0.17***	-0.07**	-0.06*	-0.11*	-0.07	-0.07**	-0.07**	-0.15**	-0.05
	(0.01)	(0.01)	(0.04)	(0.03)	(0.03)	(0.02)	(0.06)	(0.05)	(0.03)	(0.02)	(0.05)	(0.05)
Int on AL	-0.12***	-0.13***	0.00	-0.14**					-0.04†	-0.06***	0.01	-0.06
	(0.02)	(0.02)	(0.08)	(0.05)					(0.02)	(0.02)	(0.07)	(0.05)
Slope on AL	0.00	0.01*	0.01	0.01					0.00	0.00	0.02	0.00
	(0.00)	(0.00)	(0.02)	(0.01)					(0.00)	(0.00)	(0.01)	(0.01)
Int on ed					0.18***	0.17***	0.22***	0.25***	0.18***	0.17***	0.22***	0.26***
					(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)
Slope on ed					0.00	0.00	0.00	-0.01	0.00	0.00	0.00	-0.01*
					(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Int on South	-0.01	-0.01	-0.02	-0.04**	0.00	-0.01	-0.02	-0.01				
	(0.01)	(0.01)	(0.02	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)				
Slope on South	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00				
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)				

Note: WM = White men, WW = White women, BM = Black men, BW = Black women; ***p <0.001 **p<0.01 *p<0.05 †p<0.10

CHAPTER 5. CONCLUSION

As I have outlined throughout this dissertation, the aging population of the United States and around the world has initiated efforts from governments, policymakers, international agencies, and public health officials to understand the consequences of chronic illness of old age, namely Alzheimer's disease and related dementias (ADRDs) (Nichols et al. 2022; Prince et al. 2015; Rajan et al. 2021; U.S. Census Bureau 2021). The global public health concern around ADRDs—the major cause of cognitive disability in aging populations—has led to both domestic and international initiatives to gather population data to better understanding the etiology of dementia in high-, middle-, and low-income country contexts (Gross et al. 2023; Langa et al. 2020; Nichols et al. 2022). As a leader in gathering and disseminating data and reports on aging populations in the United States and abroad, the HRS's cognition assessment (Telephone Interview for Cognitive Status, or TICS) and Aging, Demographics, and Memory Study (ADAMS) paved the way for the establishment of the Harmonized Cognitive Assessment Protocol (HCAP) as an international resource to make cross-national comparisons of the causes, consequences, and trends in dementia around the world (Gross et al. 2023; Langa et al. 2005, 2020). The HCAP has been developed and implemented in international partner studies of the HRS, which include six countries (China, England, India, Mexico, South Africa, and U.S.) with the intention of providing valid and accurate cognition assessments in diverse cultural, educational, social, economic, and political contexts (Gross et al. 2023; Langa et al. 2020). Altogether, the HRS has created and made available invaluable resources in the study of cognitive aging for scholars and practitioners across a wide range of disciplines.

Specifically in the U.S. context, social scientists have made a concerted effort to explain race and gender differences in ADRDs by identifying modifiable risk factors, or the social,

contextual, and behavioral factors that can be controlled or changed and potentially lead to ADRD prevention and reduction (Livingston et al. 2020; Nianogo et al. 2022). Because race- and gender- based cognition disparities are not, by definition, essential to socially constructed social statuses, the principal drivers of disparities are the larger structural factors that produce these disparities (Braveman et al. 2011; Braveman and Gottlieb 2014; Williams 2012; Williams and Mohammed 2013; Williams and Sternthal 2010). Scholars frequently utilize the HRS's TICS assessment to estimate group differences and identify the influence that modifiable risk factors have on cognition. This dissertation extends the effort of HRS by asking if and to what extent bias in the form of measurement non-invariance in the HRS's cognition assessment contributed to disparities in cognition across four intersectional race-gender groups. Research using the HRS's TICS assessment infrequently demonstrates whether cognition is equivalently measured across race and gender groups, and even more seldomly demonstrated at the intersection of race and gender. By centering measurement invariance in cognition assessments while utilizing an intersectional lens, I was able to uncover bias in the measurement of cognition that would have otherwise gone unnoticed if race and gender were considered separately.

These analyses and findings add to the national and global efforts of creating and critically assessing how well cognition batteries, like the HRS's widely-used cognitive assessments, capture cognition (Gross et al. 2023; Langa et al. 2020). Taken together, my findings suggest that some proportion of the disparities in cognition across intersectional race-gender groups is due to biased items in the measurement tool itself, which translate into significant differences in cognition estimates across intersectional race-gender groups. My analyses are far from an indictment of the HRS and its cognition assessment, but instead work

alongside the HRS's efforts to better understand how measurement models perform in diverse populations and the subsequent cognition estimates those measurement models produce.

Summary of Key Findings

In Chapter 2, I computed the proportion of bias in the measurement of cognition across four intersectional race-gender groups (Black women, Black men, White women, White men) in a cross-sectional sample of adults 50 years and older. The important contribution of this paper was the utilization of an intersectional lens, which informed the hypotheses and revealed that the bias in the cognition instrument superficially lowered the average cognition of Black and White women. When I adjusted the measurement model so that cognition was measured equivalently across the four intersectional race-gender groups, Black women and White women had significantly higher cognition than their male counterparts. Without this adjustment, I would have come to the incorrect conclusion that cognition disparities are predominantly drawn down race lines, with little differentiation by gender (e.g., Yang et al. 2023). These results imply that measurement invariance must be taken into account in cross-sectional, observational studies to estimate the real differences between intersectional race-gender groups.

Chapter 3 extended the findings of Chapter 2 by determining the amount of bias over age trajectories of cognition across the four intersectional race-gender groups. The current literature has not studied the amount of bias due to measurement non-invariance in longitudinal studies of cognition. Using a sample of adults aged 65 and over, I compared an unadjusted, biased model with the adjusted, unbiased model established in Chapter 2 to quantify the amount of bias the cognition assessment introduced in longitudinal models of cognition. While the results were not as dramatic as the results in Chapter 2, the adjusted, unbiased model showed a clear reduction in

the gap between Black women and White men by approximately 50%. An additional contribution of this chapter was hypothesizing around the covariance between the slope and intercept, an often-overlooked coefficient in latent growth model output, across the four groups to measure the amount of variation—or uncertainty—in trajectories. I found that White women and White men had statistically significant, negative covariances between the intercept and slopes, while Black women and men had coefficients that were not statistically different from zero. This implies that the downward trajectories for Black women and men are varied, leading to greater individual-level uncertainty and greater population-level heterogeneity.

The last chapter modeled adjusted, unbiased trajectories of cognition over age, controlling for social and contextual variables that research has found to influence cognition, which included allostatic load, education, and number of years lived in the south. The results find that education had the largest association with the intercepts of the latent growth models across all four of the intersectional race-gender groups. When I controlled for education, the intercepts dropped by approximately 50%, with greater drops for Black women and men compared to White women and men, suggesting that years of education had a larger impact on the level of education. Notably, education was not associated with the slope, or change, in cognition over age. Allostatic load—a measure of elevated, sustained stress response—was negatively associated with the intercept, or level, of cognition. When I controlled for allostatic load, the intercepts for White women and men increased, suggesting that chronic, elevated stress is detrimental to cognition for White men and women. There was no such association for Black women and men. Ultimately, when all three covariates were included in the adjusted, unbiased model, the disparities between trajectories substantially diminished among Black women, Black men, and White men. The adjusting for measurement non-invariance longitudinally and

controlling for socially relevant covariates did not reduce the gap between White women and Black women, Black men, and White men.

Suggested Directions for Future Research

Throughout this dissertation, I have elucidated the importance of ensuring that cognition is measured equivalently across race-gender groups. In calculating the amount of bias that measurement non-invariance introduces, I have found substantial evidence that cross-sectional and longitudinal studies of cognition must demonstrate that the cognition measurement is free from bias to calculate true differences in cognitive ability. Utilizing an intersectional lens allowed me to uncover considerable bias in the measurement of cognition across Black women, Black men, White women, and White men, which would have gone unnoticed had race and gender been considered separately. These findings set up the potential for future studies to further investigate measurement bias in widely-used survey assessments.

Quantitative social scientist—who presumably know that race and gender bias is built into cognition assessments by virtue of being developed by White men—have the opportunity to utilize methodology that can ensure measurement non-invariance in multi-level, multi-group models. Alignment optimization (AO) (Asparouhov and Muthén 2014) within a confirmatory factor analysis (AwC) (Marsh et al. 2018) framework offers a method to ensure measurement invariance in longitudinal, multi-level, and multi-group models. The AO method corrects for measurement invariance in large, complex latent measurement models, creating a single measurement model akin to the configural invariance model (Asparouhov and Muthén 2023; Muthén and Asparouhov 2018). The AwC method extends the AO method by using the optimized (i.e., invariant) measurement model in a growth model (Lai 2021; Marsh et al. 2018),

steps similar to the second-order growth model performed in Chapters 3 and 4. As sociologists, we are aware of the potential for bias in cognition assessments, and yet, very little attention is paid to identifying and correcting for bias in the measurement of cognition. There is room for collaboration between methodologists and sociologists to use these methods with real data to better understand aging and cognition.

Additionally, I alluded to the importance of elucidating all aspects of inequality in ADRDs, including considering how variation in cognition in cognition trajectories may impact the prognosis and sense of control one has on their health in old age. Previous research has investigated the average number of years people are expected to live with an ADRD, disaggregated by race and education level (Farina et al. 2020; Garcia et al. 2021). Much like demographic scholars have done with lifespan variation and dispersion (e.g., Firebaugh et al. 2014; van Raalte et al. 2018; van Raalte and Martikainen 2014; Sasson 2016), a potentially fruitful next step is to utilize official enumerations of ADRDs and age at diagnosis to better understand variation around ADRDs across groups. The dispersion around the average years lived with an ADRD points to when in the life course people are being diagnosed and provides a potential entry point in intervening or delaying the onset of dementia. These differences in dispersion around dementia-free life expectancy may uncover more fully the disparities people face.

Overall, the intersectional framework I used opens the possibility of executing measurement invariance analyses using intersection of other social identities or in contexts outside the United States. For example, educational attainment may be an achieved status that effects answering cognition assessment items related to crystallized knowledge, or the accumulation of knowledge and facts one accumulates through school. Similarly, those who

speak English as a second language face barriers to answering questions that has little to do with their actual cognitive ability. Measurement invariance analysis across intersectional groups in international contexts may uncover non-invariance across diverse groups or provide further evidence for the validity of cognition assessments in culturally and linguistically varied contexts. These additional social statuses and contexts make conveying cognition challenging; it is important to consider these and other identities when analyzing cognition data and ensuring the measurements are unbiased.

REFERENCES

Adkins-Jackson, Paris B., Kristen M. George, Lilah M. Besser, Jinshil Hyun, Melissa Lamar, Tanisha G. Hill-Jarrett, Omonigho M. Bubu, Jason D. Flatt, Patricia C. Heyn, Ethan C. Cicero, A. Zarina Kraal, Preeti Pushpalata Zanwar, Rachel Peterson, Boeun Kim, Robert W. Turner, Jaya Viswanathan, Erin R. Kulick, Megan Zuelsdorff, Shana D. Stites, Miguel Arce Rentería, Elena Tsoy, Dominika Seblova, Ted K. S. Ng, Jennifer J. Manly, and Ganesh Babulal. 2023. "The Structural and Social Determinants of Alzheimer's Disease Related Dementias." *Alzheimer's and Dementia* 19(7):3171–85.

Agénor, Madina. 2020. "Future Directions for Incorporating Intersectionality Into Quantitative Population Health Research." *American Journal of Public Health* 110(6):803–6. doi: 10.2105/AJPH.2020.305610.

Agénor, Madina, Zinzi Bailey, Nancy Krieger, S. Bryn Austin, and Barbara R. Gottlieb. 2015. "Exploring the Cervical Cancer Screening Experiences of Black Lesbian, Bisexual, and Queer Women: The Role of Patient-Provider Communication." *Women and Health* 55(6):717–36. doi: 10.1080/03630242.2015.1039182.

Ailshire, Jennifer A., Katrina M. Walsemann, and Calley E. Fisk. 2022. "Regional Variation in U.S Dementia Trends from 2000-2012." *SSM - Population Health* 19. doi: 10.1016/j.ssmph.2022.101164.

Allen, Amani M., Yijie Wang, David H. Chae, Melisa M. Price, Wizdom Powell, Teneka C. Steed, Angela Rose Black, Firdaus S. Dhabhar, Leticia Marquez-Magaña, and Cheryl L. Woods-Giscombe. 2019. "Racial Discrimination, the Superwoman Schema, and Allostatic Load: Exploring an Integrative Stress-Coping Model among African American Women." *Annals of the New York Academy of Sciences* 1457(1):104–27. doi: 10.1111/nyas.14188.

Alzheimer's Association. 2020. "2020 Alzheimer's Disease Facts and Figures." *Alzheimer's and Dementia* 16(3):391–460. doi: 10.1002/alz.12068.

Alzheimer's Association. 2022. *More Than Normal Aging: Understanding Mild Cognitive Impairment*. Chicago, IL.

Anderson, Kathryn Freeman. 2013. "Diagnosing Discrimination: Stress from Perceived Racism and the Mental and Physical Health Effects." *Sociological Inquiry* 83(1):55–81. doi: 10.1111/J.1475-682X.2012.00433.X.

Ansell, Emily B., Kenneth Rando, Keri Tuit, Joseph Guarnaccia, and Rajita Sinha. 2012. "Cumulative Adversity and Smaller Gray Matter Volume in Medial Prefrontal, Anterior Cingulate, and Insula Regions." *Biological Psychiatry* 72(1):57–64. doi: 10.1016/j.biopsych.2011.11.022.

Asparouhov, Tihomir, and Bengt Muthén. 2014. "Multiple-Group Factor Analysis Alignment." *Structural Equation Modeling* 21(4):495–508. doi: 10.1080/10705511.2014.919210.

Asparouhov, Tihomir, and Bengt Muthén. 2023. "Multiple Group Alignment for Exploratory and Structural Equation Models." *Structural Equation Modeling* 30(2):169–91. doi: 10.1080/10705511.2022.2127100.

Avila, Justina F., Miguel Arce Rentería, Richard N. Jones, Jet M. J. Vonk, Indira Turney, Ketlyne Sol, Dominika Seblova, Franchesca Arias, Tanisha Hill-Jarrett, Shellie Anne Levy, Oanh

Meyer, Annie M. Racine, Sarah E. Tom, Rebecca J. Melrose, Kacie Deters, Luis D. Medina, Carmen I. Carrión, Mirella Díaz-Santos, De Annah R. Byrd, Anthony Chesebro, Juliet Colon, Kay C. Igwe, Benjamin Maas, Adam M. Brickman, Nicole Schupf, Richard Mayeux, and Jennifer J. Manly. 2021. "Education Differentially Contributes to Cognitive Reserve across Racial/Ethnic Groups." *Alzheimer's and Dementia* 17(1):70–80. doi: 10.1002/alz.12176.

Avila, Justina F., Miguel Arce Rentería, Katie Witkiewitz, Steven P. Verney, Jet M. J. Vonk, and Jennifer J. Manly. 2020. "Measurement Invariance of Neuropsychological Measures of Cognitive Aging Across Race/Ethnicity by Sex/Gender Groups." *Neuropsychology* 34(1):3–14. doi: 10.1037/neu0000584.

Bailey, Zinzi D., Nancy Krieger, Madina Agénor, Jasmine Graves, Natalia Linos, and Mary T. Bassett. 2017. "Structural Racism and Health Inequities in the USA: Evidence and Interventions." *The Lancet* 389(10077):1453–63. doi: https://doi.org/10.1016/S0140-6736(17)30569-X.

Barnes, Lisa L., Futoshi Yumoto, Ana Capuano, Robert S. Wilson, David A. Bennett, and Rochelle E. Tractenberg. 2015. "Examination of the Factor Structure of a Global Cognitive Function Battery across Race and Time." *Journal of the International Neuropsychological Society* 22(1):66–75. doi: 10.1017/S1355617715001113.

Bauer, Greta R. 2014. "Incorporating Intersectionality Theory into Population Health Research Methodology: Challenges and the Potential to Advance Health Equity." *Social Science and Medicine* 110:10–17. doi: 10.1016/j.socscimed.2014.03.022.

Bauer, Greta R., Siobhan M. Churchill, Mayuri Mahendran, Chantel Walwyn, Daniel Lizotte, and Alma Angelica Villa-Rueda. 2021. "Intersectionality in Quantitative Research: A Systematic Review of Its Emergence and Applications of Theory and Methods." *SSM - Population Health* 14. doi: 10.1016/j.ssmph.2021.100798.

Bauer, Greta R., and Ayden I. Scheim. 2019. "Advancing Quantitative Intersectionality Research Methods: Intracategorical and Intercategorical Approaches to Shared and Differential Constructs." *Social Science and Medicine* 226:260–62.

Beal, Frances M. 2008. "Double Jeopardy: To Be Black and Female." *Meridians: Feminism, Race, Transnationalism* 8(2):166–76.

Beech, Bettina M., Chandra Ford, Roland J. Thorpe, Marino A. Bruce, and Keith C. Norris. 2021. "Poverty, Racism, and the Public Health Crisis in America." *Frontiers in Public Health* 9. doi: 10.3389/fpubh.2021.699049.

Berkman, Lisa F., and M. Maria Glymour. 2006. "How Society Shapes Aging: The Centrality of Variability." *Daedalus* 135(1):105–14.

Blankson, A. Nayena, and John J. McArdle. 2015. "Measurement Invariance of Cognitive Abilities Across Ethnicity, Gender, and Time among Older Americans." *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 70(3):386–97. doi: 10.1093/geronb/gbt106.

Boen, Courtney E., Y. Claire Yang, Allison E. Aiello, Alexis C. Dennis, Kathleen Mullan Harris, Dayoon Kwon, Dan Iel, and W. Belsky. 2023. "Patterns and Life Course Determinants of

Black–White Disparities in Biological Age Acceleration: A Decomposition Analysis." *Demography*. doi: 10.1215/00703370-11057546.

Bohk-Ewald, Christina, Marcus Ebeling, and Roland Rau. 2017. "Lifespan Disparity as an Additional Indicator for Evaluating Mortality Forecasts." *Demography* 54(4):1559–77. doi: 10.1007/s13524-017-0584-0.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York, NY: John Wiley and Sons, Inc.

Bowleg, Lisa. 2008. "When Black + Lesbian + Woman ≠ Black Lesbian Woman: The Methodological Challenges of Qualitative and Quantitative Intersectionality Research." *Sex Roles* 59(5–6):312–25.

Bowleg, Lisa. 2012. "The Problem With the Phrase Women and Minorities: Intersectionality—an Important Theoretical Framework for Public Health." *American Journal of Public Health* 102(7):1267–73.

Braveman, Paula A., Catherine Cubbin, Susan Egerter, David R. Williams, and Elsie Pamuk. 2010. "Socioeconomic Disparities in Health in the United States: What the Patterns Tell Us." *Am J Public Health* 100(S1):S186–96. doi: 10.2105/ajph.2009.166082.

Braveman, Paula A., Shiriki Kumanyika, Jonathan Fielding, Thomas LaVeist, Luisa N. Borrell, Ron Manderscheid, and Adewale Troutman. 2011. "Health Disparities and Health Equity: The Issue Is Justice." *American Journal of Public Health* 101(S1):S149–55.

Braveman, Paula, and Laura Gottlieb. 2014. *The Social Determinants of Health: It's Time to Consider the Causes of the Causes*. Vol. 129.

Byrne, Barbara M. 2011. *Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming*. Taylor & Francis Group.

Byrne, Barbara M., and Gail Crombie. 2003. "Modeling and Testing Change: An Introduction to the Latent Growth Curve Model." *Understanding Statistics* 2(3):177–203.

Byrne, Kaileigh A., and Reza Ghaiumy Anaraky. 2022. "Identifying Racial and Rural Disparities of Cognitive Functioning among Older Adults: The Role of Social Isolation and Social Technology Use." *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 77(10):1779–90. doi: 10.1093/geronb/gbac055.

Carbado, Devon W. 2013. "Colorblind Intersectionality." *Signs: Journal of Women in Culture and Society* 38(4):811–45.

Carbado, Devon W., and Daria Roithmayr. 2014. "Critical Race Theory Meets Social Science." *Annual Review of Law and Social Science* 10:149–67. doi: 10.1146/annurev-lawsocsci-110413-030928.

Center for Disease Control and Prevention. 2023. "LDL and HDL Cholesterol and Triglycerides." *National Center for Chronic Disease Prevention and Health Promotion, Division for Heart Disease and Stroke Prevention*.

Chan, David. 1998. "The Conceptualization and Analysis of Change Over Time: An Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS)

and Multiple Indicator Latent Growth Modeling (MLGM)." *Https://Doi.Org/10.1177/109442819814004* 1(4):421–83. doi: 10.1177/109442819814004.

Chen, Cynthia, and Julie M. Zissimopoulos. 2018. "Racial and Ethnic Differences in Trends in Dementia Prevalence and Risk Factors in the United States." *Alzheimer's and Dementia: Translational Research and Clinical Interventions* 4:510–20. doi: 10.1016/j.trci.2018.08.009.

Chen, Fang Fange. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling: A Multidisciplinary Journal* 14(3):464–504.

Chen, Ruijia, Jennifer Weuve, Supriya Misra, Adolfo Cuevas, Laura D. Kubzansky, and David R. Williams. 2022. "Racial Disparities in Cognitive Function Among Middle-Aged and Older Adults: The Roles of Cumulative Stress Exposures Across the Life Course." *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 77(2):357–64. doi: 10.1093/gerona/glab099.

Chêne, Geneviève, Alexa Beiser, Rhoda Au, Sarah R. Preis, Philip A. Wolf, Carole Dufouil, and Sudha Seshadri. 2015. "Gender and Incidence of Dementia in the Framingham Heart Study from Mid-Adult Life." *Alzheimer's and Dementia* 11(3):310–20. doi: 10.1016/j.jalz.2013.10.005.

Cheung, Gordon W., and Roger B. Rensvold. 1999. "Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method." *Journal of Management* 25(1):1–27. doi: 10.1016/S0149-2063(99)80001-4.

Cheung, Gordon W., and Roger B. Rensvold. 2002. "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance." *Structural Equation Modeling* 9(2):233–55. doi: 10.1207/S15328007SEM0902_5.

Cho, Sumi, Kimberlé Williams Crenshaw, and Leslie McCall. 2013. "Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis." *Signs: Journal of Women in Culture and Society* 38(4):785–810.

Chou, Chih Ping, Peter M. Bentler, and Mary Ann Pentz. 1998. "Comparisons of Two Statistical Approaches to Study Growth Curves: The Multilevel Model and the Latent Curve Analysis." *Structural Equation Modeling* 5(3):247–66. doi: 10.1080/10705519809540104.

Christensen, Kaare, Gabriele Doblhammer, Roland Rau, and James W. Vaupel. 2009. *Ageing Populations: The Challenges Ahead*. Vol. 374.

Cintron, Dakota W., Camilla Calmasini, Lisa L. Barnes, Dan M. Mungas, Rachel A. Whitmer, Chloe W. Eng, Paola Gilsanz, Kristen M. George, Rachel L. Peterson, and M. Maria Glymour. 2023. "Evaluating Interpersonal Discrimination and Depressive Symptoms as Partial Mediators of the Effects of Education on Cognition: Evidence from the Study of Healthy Aging in African Americans (STAR)." *Alzheimer's and Dementia*. doi: 10.1002/alz.12957.

Clark, D. Angus, and M. Brent Donnellan. 2021. "What If Apples Become Oranges? A Primer on Measurement Invariance in Repeated Measures Research." Pp. 837–54 in *The Handbook of Personality Dynamics and Processes*. Elsevier.

Collins, Patricia Hill. 2015. “Intersectionality’s Definitional Dilemmas.” *Annual Review of Sociology* 41:1–20.

Coogan, Patricia, Karin Schon, Shanshan Li, Yvette Cozier, Traci Bethea, and Lynn Rosenberg. 2020. “Experiences of Racism and Subjective Cognitive Function in African American Women.” *Alzheimer’s and Dementia: Diagnosis, Assessment and Disease Monitoring* 12(1). doi: 10.1002/dad2.12067.

Crenshaw, Kimberle. 1989. “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics.” *University of Chicago Legal Forum* 1989(1):139–67.

Crenshaw, Kimberle. 1991. *Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color*. Vol. 43.

Crenshaw, Kimberle. 2017. *On Intersectionality: Essential Writings*. New York, NY: The New Press.

Crimmins, Eileen, Jessica Faul, Jung Ki Kim, and David Weir. 2020. *Documentation of DBS Blood-Based Biomarkers in the 2016 Health and Retirement Study*. Ann Arbor, MI.

Crimmins, Eileen, Heidi Guyer, Kenneth Langa, Mary Beth Ofstedal, Robert Wallace, and David Weir. 2008. *Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study*. Ann Arbor, MI.

Crimmins, Eileen M. 2020. “Social Hallmarks of Aging: Suggestions for Geroscience Research.” *Ageing Research Reviews* 63. doi: https://doi.org/10.1016/j.arr.2020.101136.

Crimmins, Eileen M., and Yasuhiko Saito. 2001. “Trends in Healthy Life Expectancy in the United States, 1970–1990: Gender, Racial, and Educational Differences.” *Social Science & Medicine* 52(11):1629–41. doi: 10.1016/S0277-9536(00)00273-2.

Crimmins, Eileen M., Yasuhiko Saito, Jung Ki Kim, Yuan S. Zhang, Isaac Sasson, and Mark D. Hayward. 2018. “Educational Differences in the Prevalence of Dementia and Life Expectancy with Dementia: Changes from 2000 to 2010.” *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 73:S20–28.

Crimmins, Eileen M., and Yuan S. Zhang. 2019. “Aging Populations, Mortality, and Life Expectancy.” *Annual Review of Sociology* 45(1):annurev-soc-073117-041351. doi: 10.1146/annurev-soc-073117-041351.

D’Amico, Danielle, Maya E. Amestoy, and Alexandra J. Fiocco. 2020. “The Association between Allostatic Load and Cognitive Function: A Systematic and Meta-Analytic Review.” *Psychoneuroendocrinology* 121. doi: 10.1016/j.psyneuen.2020.104849.

Diwadkar, Vaibhav A. 2016. “Epigenetics, Stress, and Their Potential Impact on Brain Network Function.” Pp. 127–35 in *Stress: Concepts, Cognition, Emotion, and Behavior*, edited by G. Fink. San Diego: Academic Press.

Dufouil, Carole, Alexa Beiser, Geneviève Chêne, and Sudha Seshadri. 2018. “Are Trends in Dementia Incidence Associated with Compression in Morbidity? Evidence from the Framingham Heart Study.” *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 73:S65–72.

Enders, Craig K., and Deborah L. Bandalos. 2001. “The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models.” *Structural Equation Modeling* 8(3):430–57. doi: 10.1207/S15328007SEM0803_5.

Engelman, Michal, Vladimir Canudas-Romo, and Emily M. Agree. 2010. “The Implications of Increased Survivorship for Mortality Variation in Aging Populations.” *Population and Development Review* 36(3):511–39. doi: 10.1111/j.1728-4457.2010.00344.x.

Farina, Mateo P., Mark D. Hayward, Jung Ki Kim, and Eileen M. Crimmins. 2020. “Racial and Educational Disparities in Dementia and Dementia-Free Life Expectancy.” *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 75(7):E105–12. doi: 10.1093/geronb/gbz046.

Faul, Jessica D., Erin B. Ware, Mohammed U. Kabeto, Jonah Fisher, and Ken M. Langa. 2021. “The Effect of Childhood Socioeconomic Position and Social Mobility on Cognitive Function and Change among Older Adults: A Comparison between the United States and England.” *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 76:S51–63. doi: 10.1093/geronb/gbaa138.

Ferrer, Emilio, Nekane Balluerka, and Keith F. Widaman. 2008. “Factorial Invariance and Specification of Second-Order Latent Growth Models.” *Methodology* 4(1):22–36. doi: 10.1027/1614-2241.4.1.22.

Firebaugh, Glenn, Francesco Acciai, Aggie J. Noah, Christopher Prather, and Claudia Nau. 2014. “Why Lifespans Are More Variable Among Blacks Than Among Whites in the United States.” *Demography* 51(6):2025–45. doi: 10.1007/s13524-014-0345-2.

Fisher, Gwenith G., Dorey S. Chaffee, Lois E. Tetrick, Deana B. Davalos, and Guy G. Potter. 2017. “Cognitive Functioning, Aging, and Work: A Review and Recommendations for Research and Practice.” *Journal of Occupational Health Psychology* 22(3):314–36. doi: 10.1037/ocp0000086.

Fitzpatrick, Annette L., Lewis H. Kuller, Diane G. Ives, Oscar L. Lopez, William Jagust, John C. S. Breitner, Beverly Jones, Constantine Lyketsos, and Corinne Dulberg. 2004. “Incidence and Prevalence of Dementia in the Cardiovascular Health Study.” *Journal of the American Geriatrics Society* 52(2):195–204. doi: 10.1111/j.1532-5415.2004.52058.x.

Fletcher, Jason M., Hans Schwarz, Michal Engelman, Norman J. Johnson, Jahn Hakes, and Alberto Palloni. 2023. “Understanding Geographic Disparities in Mortality.” *Demography* 60(2):351–77. doi: 10.1215/00703370-10609710.

Fries, James F. 1980. “Aging, Natural Death, and the Compression of Morbidity.” *The New England Journal of Medicine* 303(3):130–35.

Fries, James F. 2005. “The Compression of Morbidity.” *The Milbank Quarterly* 83(4):801–23.

Fujishiro, Kaori, Leslie A. MacDonald, Michael Crowe, Leslie A. McClure, Virginia J. Howard, and Virginia G. Wadley. 2019. “The Role of Occupation in Explaining Cognitive Functioning in Later Life: Education and Occupational Complexity in a U.S. National Sample of Black and White Men and Women.” *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 74(7):1189–99. doi: 10.1093/geronb/gbx112.

Ganzel, Barbara L., Pamela A. Morris, and Elaine Wethington. 2010. “Allostasis and the Human Brain: Integrating Models of Stress From the Social and Life Sciences.” *Psychological Review* 117(1):134–74. doi: 10.1037/a0017773.

Garcia, Marc A., Brian Downer, Chi Tsun Chiu, Joseph L. Saenz, Kasim Ortiz, and Rebeca Wong. 2021. “Educational Benefits and Cognitive Health Life Expectancies: Racial/Ethnic, Nativity, and Gender Disparities.” *Gerontologist* 61(3):330–40. doi: 10.1093/geront/gnaa112.

Geiser, Christian, Brian T. Keller, and Ginger Lockhart. 2013. “First- Versus Second-Order Latent Growth Curve Models: Some Insights From Latent State-Trait Theory.” *Structural Equation Modeling: A Multidisciplinary Journal* 20(3):479–503. doi: 10.1080/10705511.2013.797832.

Gilsanz, Paola, Elizabeth Rose Mayeda, M. Maria Glymour, Charles P. Quesenberry, and Rachel A. Whitmer. 2017. “Association between Birth in a High Stroke Mortality State, Race, and Risk of Dementia.” *JAMA Neurology* 74(9):1056–62. doi: 10.1001/jamaneurol.2017.1553.

Glymour, M. Maria, and Jennifer J. Manly. 2008. “Lifecourse Social Conditions and Racial and Ethnic Patterns of Cognitive Aging.” *Neuropsychology Review* 18(3 SPEC. ISS.):223–54.

Gross, Alden L., Chihua Li, Emily M. Briceño, Miguel Arce Rentería, Richard N. Jones, Kenneth M. Langa, Jennifer J. Manly, Emma Nichols, David Weir, Rebeca Wong, Lisa Berkman, Jinkook Lee, and Lindsay C. Kobayashi. 2023. “Harmonisation of Later-Life Cognitive Function across National Contexts: Results from the Harmonized Cognitive Assessment Protocols.” *Lancet Healthy Longevity* 4:e573–83.

Hale, Jo Mhairi, Daniel C. Schneider, Jutta Gampe, Neil K. Mehta, and Mikko Myrskylä. 2020. “Trends in the Risk of Cognitive Impairment in the United States, 1996-2014.” *Epidemiology* 31(5):745–54. doi: 10.1097/EDE.0000000000001219.

Hale, Jo Mhairi, Daniel C. Schneider, Neil K. Mehta, and Mikko Myrskylä. 2020. “Cognitive Impairment in the U.S.: Lifetime Risk, Age at Onset, and Years Impaired.” *SSM - Population Health* 11. doi: 10.1016/j.ssmph.2020.100577.

Harari, Lexi, and Chioun Lee. 2021. “Intersectionality in Quantitative Health Disparities Research: A Systematic Review of Challenges and Limitations in Empirical Studies.” *Social Science and Medicine* 277:1–11. doi: 10.1016/j.socscimed.2021.113876.

Harnois, Catherine E., and Mosi Ifatunji. 2011. “Gendered Measures, Gendered Models: Toward an Intersectional Analysis of Interpersonal Racial Discrimination.” *Ethnic and Racial Studies* 34(6):1006–28. doi: 10.1080/01419870.2010.516836.

Hayward, Mark D., Mateo P. Farina, Yuan S. Zhang, Jung Ki Kim, and Eileen M. Crimmins. 2021. “The Importance of Improving Educational Attainment for Dementia Prevalence Trends From 2000 to 2014, Among Older Non-Hispanic Black and White Americans.” *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 76(9):1870–79. doi: 10.1093/geronb/gbab015.

Health and Retirement Study. 2022. “RAND HRS Longitudinal File 2018 (V2) Public Use Dataset.”

Hebert, Liesi E., Paul A. Scherr, Judith J. McCann, Laurel A. Beckett, and Denis A. Evans. 2001. "Is the Risk of Developing Alzheimer's Disease Greater for Women than for Men?" *American Journal of Epidemiology* 153(2):132–36. doi: 10.1093/aje/153.2.132.

Heeringa, Steven G., Gwenith G. Fisher, Michael Hurd, Kenneth M. Langa, Mary Beth Ofstedal, Brenda L. Plassman, Willard L. Rodgers, and David R. Weir. 2009. *Aging, Demographics and Memory Study (ADAMS): Sample Design, Weighting and Analysis for ADAMS*. Ann Arbor.

Herzog, A. Regula, and Robert B. Wallace. 1997. "Measures of Cognitive Functioning in the AHEAD Study." *The Journals of Gerontology: Series B* 52B(Special_Issue):37–48. doi: 10.1093/geronb/52B.Special_Issue.37.

Horn, John L., and J. J. McArdle. 1992. "A Practical and Theoretical Guide to Measurement Invariance in Aging Research." *Quantitative Topics in Research in Aging* 18(3):117–44.

Howard, George, and Virginia J. Howard. 2020. "Twenty Years of Progress Toward Understanding the Stroke Belt." *Stroke* 51(3):742–50. doi: 10.1161/STROKEAHA.119.024155.

Inker, Lesley A., and Andrew S. Levey. 2018. "Assessment of Kidney Function in Acute and Chronic Settings." Pp. 26-32.e1 in *National Kidney Foundation's Primer on Kidney Diseases (Seventh Edition)*, edited by S. J. Gilbert and D. E. Weiner. Philadelphia: Elsevier.

Jackson, Fleda Mask, Carol Rowland Hogue, and Mona Taylor Phillips. 2005. "The Development of a Race and Gender-Specific Stress Measure for African-American Women on JSTOR." *Ethnicity & Disease* 15(4):594–600.

Karlamangla, Arun S., Burton H. Singer, Bruce S. McEwen, John W. Rowe, and Teresa E. Seeman. 2002. *Allostatic Load as a Predictor of Functional Decline MacArthur Studies of Successful Aging*. Vol. 55.

Kemp, Blakelee R., Jacob M. Grumbach, and Jennifer Karas Montez. 2022. "U.S. State Policy Contexts and Physical Health among Midlife Adults." *Socius* 8. doi: 10.1177/23780231221091324.

King, Deborah K. 1988. "Multiple Jeopardy, Multiple Consciousness: The Context of a Black Feminist Ideology." *Signs* 14(1):42–72.

Kline, Rex. 2015. *Principles and Practice of Structural Equation Modeling*. 4th ed. New York, NY: Guilford Press.

Knighton, Joi Sheree', Jardin Dogan, Candice Hargons, and Danelle Stevens-Watkins. 2022. "Superwoman Schema: A Context for Understanding Psychological Distress among Middle-Class African American Women Who Perceive Racial Microaggressions." *Ethnicity and Health* 27(4):946–62. doi: 10.1080/13557858.2020.1818695.

Lai, Keke, and Samuel B. Green. 2016. "The Problem with Having Two Watches: Assessment of Fit When RMSEA and CFI Disagree." *Multivariate Behavioral Research* 51(2–3):220–39. doi: 10.1080/00273171.2015.1134306.

Lai, Mark H. C. 2021. "Adjusting for Measurement Noninvariance with Alignment in Growth Modeling." *Multivariate Behavioral Research*. doi: 10.1080/00273171.2021.1941730.

Langa, Kenneth M. 2015. “Is the Risk of Alzheimer’s Disease and Dementia Declining?” in *Alzheimer’s Research and Therapy*. Vol. 7. BioMed Central Ltd.

Langa, Kenneth M., Eric B. Larson, Jason H. Karlawish, David M. Cutler, Mohammed U. Kabeto, Scott Y. Kim, and Allison B. Rosen. 2008. “Trends in the Prevalence and Mortality of Cognitive Impairment in the United States: Is There Evidence of a Compression of Cognitive Morbidity?” *Alzheimer’s and Dementia* 4(2):134–44. doi: 10.1016/j.jalz.2008.01.001.

Langa, Kenneth M., Brenda L. Plassman, Robert B. Wallace, A. Regula Herzog, Steven G. Heeringa, Mary Beth Ofstedal, James R. Burke, Gwenith G. Fisher, Nancy H. Fultz, Michael D. Hurd, Guy G. Potter, Willard L. Rodgers, David C. Steffens, David R. Weir, and Robert J. Willis. 2005. “The Aging, Demographics, and Memory Study: Study Design and Methods.” *Neuroepidemiology* 25(4):181–91. doi: 10.1159/000087448.

Langa, Kenneth M., Lindsay H. Ryan, Ryan J. McCammon, Richard N. Jones, Jennifer J. Manly, Deborah A. Levine, Amanda Sonnega, Madeline Farron, and David R. Weir. 2020. “The Health and Retirement Study Harmonized Cognitive Assessment Protocol Project: Study Design and Methods.” *Neuroepidemiology* 54(1):64–74. doi: 10.1159/000503004.

Leggett, Amanda, Philippa Clarke, Kara Zivin, Ryan J. McCammon, Michael R. Elliott, and Kenneth M. Langa. 2019. “Recent Improvements in Cognitive Functioning among Older U.S. Adults: How Much Does Increasing Educational Attainment Explain?” *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 74(3):536–42. doi: 10.1093/geronb/gbw210.

Leitgöb, Heinz, Daniel Seddig, Tihomir Asparouhov, Dorothée Behr, Eldad Davidov, Kim De Roover, Suzanne Jak, Katharina Meitinger, Natalja Menold, Bengt Muthén, Maksim Rudnev, Peter Schmidt, and Rens van de Schoot. 2023. “Measurement Invariance in the Social Sciences: Historical Development, Methodological Challenges, State of the Art, and Future Perspectives.” *Social Science Research* 110. doi: 10.1016/j.ssresearch.2022.102805.

Lenehan, Megan Elizabeth, Mathew James Summers, Nichole Louise Saunders, Jeffery Joseph Summers, and James C. Vickers. 2015. “Relationship between Education and Age-Related Cognitive Decline: A Review of Recent Research.” *Psychogeriatrics* 15(2):154–62.

Letang, Sarah K., Shayne S. H. Lin, Patricia A. Parmelee, and Ian M. McDonough. 2021. “Ethnoracial Disparities in Cognition Are Associated with Multiple Socioeconomic Status-Stress Pathways.” *Cognitive Research: Principles and Implications* 6(1). doi: 10.1186/s41235-021-00329-7.

Levine, Deborah A., Alden L. Gross, Emily M. Briceño, Nicholas Tilton, Bruno J. Giordani, Jeremy B. Sussman, Rodney A. Hayward, James F. Burke, Stephanie Hingtgen, Mitchell S. V. Elkind, Jennifer J. Manly, Rebecca F. Gottesman, Darrell J. Gaskin, Stephen Sidney, Ralph L. Sacco, Sarah E. Tom, Clinton B. Wright, Kristine Yaffe, and Andrzej T. Galecki. 2021. “Sex Differences in Cognitive Decline among US Adults.” *JAMA Network Open* 4(2). doi: 10.1001/jamanetworkopen.2021.0169.

Liu, Chelsea, Audrey R. Murchland, Tyler J. VanderWeele, and Deborah Blacker. 2022. “Eliminating Racial Disparities in Dementia Risk by Equalizing Education Quality: A Sensitivity Analysis.” *Social Science and Medicine* 312. doi: 10.1016/j.socscimed.2022.115347.

Liu, Sze Yan, M. Maria Glymour, Laura B. Zahodne, Christopher Weiss, and Jennifer J. Manly. 2015. "Role of Place in Explaining Racial Heterogeneity in Cognitive Outcomes among Older Adults." *Journal of the International Neuropsychological Society* 21(9):677–87. doi: 10.1017/S1355617715000806.

Livingston, Gill, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, Sergi G. Costafreda, Amit Dias, Nick Fox, Laura N. Gitlin, Robert Howard, Helen C. Kales, Mika Kivimäki, Eric B. Larson, Adesola Ogunniyi, Vasiliki Orgeta, Karen Ritchie, Kenneth Rockwood, Elizabeth L. Sampson, Quincy Samus, Lon S. Schneider, Geir Selbæk, Linda Teri, and Naaheed Mukadam. 2020. "Dementia Prevention, Intervention, and Care: 2020 Report of the Lancet Commission." *The Lancet* 396(10248):413–46.

Lövdén, Martin, Laura Fratiglioni, M. Maria Glymour, Ulman Lindenberger, and Elliot M. Tucker-Drob. 2020. "Education and Cognitive Functioning Across the Life Span." *Psychological Science in the Public Interest* 21(1):6–41. doi: 10.1177/1529100620920576.

Maccora, Janet, Ruth Peters, and Kaarin J. Anstey. 2020. "What Does (Low) Education Mean in Terms of Dementia Risk? A Systematic Review and Meta-Analysis Highlighting Inconsistency in Measuring and Operationalising Education." *SSM - Population Health* 12.

Marsh, Herbert W., Jiesi Guo, Philip D. Parker, Benjamin Nagengast, Tihomir Asparouhov, Bengt Muthén, and Theresa Dicke. 2018. "What to Do When Scalar Invariance Fails: The Extended Alignment Method for Multi-Group Factor Analysis Comparison of Latent Means across Many Groups." *Psychological Methods* 23(3):524–45. doi: 10.1037/met0000113.

Martin, Linda G., Robert F. Schoeni, and Patricia M. Andreski. 2010. "Trends in Health of Older Adults in the United States: Past, Present, Future." *Demography* 47:17–40.

McArdle, John J. 1988. "Dynamic but Structural Equation Modeling of Repeated Measures Data." Pp. 561–614 in *Handbook of Multivariate Experimental Psychology*, edited by J. R. Nesselroade and R. B. Cattell. Boston, MA: Springer US.

McArdle, John J. 2011. "Longitudinal Dynamic Analyses of Cognition in the Health and Retirement Study Panel." *AStA Advances in Statistical Analysis* 95(4):453–80. doi: 10.1007/s10182-011-0168-z.

McArdle, John J., and David Epstein. 1987. "Latent Growth Curves within Developmental Structural Equation Models." *Child Development* 58:110–33.

McArdle, John J., Gwenith G. Fisher, and Kelly M. Kadlec. 2007. "Latent Variable Analyses of Age Trends of Cognition in the Health and Retirement Study, 1992-2004." *Psychology and Aging* 22(3):525–45. doi: 10.1037/0882-7974.22.3.525.

McCammon, Ryan J., Gwenith G. Fisher, Halimah Hassan, Jessica D. Faul, Wilard L. Rodgers, and David R. Weir. 2022. *Health and Retirement Study Imputation of Cognitive Functioning Measures: 1992 – 2018*. Ann Arbor, MI.

McDonough, Ian M., De Annah R. Byrd, and Shinae L. Choi. 2023. "Resilience Resources May Buffer Some Middle-Aged and Older Black Americans from Memory Decline despite Experiencing Discrimination." *Social Science and Medicine* 316. doi: 10.1016/j.socscimed.2022.114998.

McDonough, Ian M., Shameka L. Cody, Erin R. Harrell, Stephanie L. Garrett, and Taylor E. Popp. 2022. “Cognitive Differences across Ethnoracial Category, Socioeconomic Status across the Alzheimer’s Disease Spectrum: Can an Ability Discrepancy Score Level the Playing Field?” *Memory & Cognition*. doi: 10.3758/s13421-022-01304-3.

McEwen, B. S. 2016. *Central Role of the Brain in Stress and Adaptation: Allostasis, Biological Embedding, and Cumulative Change*. Elsevier Inc.

McEwen, Bruce S. 2016. “Central Role of the Brain in Stress and Adaptation: Allostasis, Biological Embedding, and Cumulative Change.” Pp. 39–55 in *Stress: Concepts, Cognition, Emotion, and Behavior*, edited by G. Fink. San Diego: Academic Press.

McEwen, Bruce S., and Peter J. Gianaros. 2010. “Central Role of the Brain in Stress and Adaptation: Links to Socioeconomic Status, Health, and Disease.” *Annals of the New York Academy of Sciences* 1186:190–222. doi: 10.1111/j.1749-6632.2009.05331.x.

McEwen, Bruce S., and Teresa Seeman. 1999. “Protective and Damaging Effects of Mediators of Stress: Elaborating and Testing the Concepts of Allostasis and Allostatic Load.” *Ann N Y Acad Sci* 896(1):30–47. doi: 10.1111/j.1749-6632.1999.tb08103.x.

McEwen, Bruce S., and Eliot Stellar. 1993. “Stress and the Individual Mechanisms Leading to Disease.” *Arch Intern Med* 153:2093–2101.

Meredith, William. 1993. “Measurement Invariance, Factor Analysis, and Factorial Invariance.” *Psychometrika* 58(4):525–43.

Meredith, William, and Jeanne A. Teresi. 2006. *An Essay on Measurement and Factorial Invariance*. Vol. 44.

Meredith, William, and John Tisak. 1990. “Latent Curve Analysis.” *Psychometrika* 55(1):107–22.

Millsap, Roger E., and Jenn Yun-Tein. 2004. “Assessing Factorial Invariance in Ordered-Categorical Measures.” *Multivariate Behavioral Research* 39(3):479–515.

Montez, Jennifer Karas, and Mateo P. Farina. 2021. “Do Liberal U.S. State Policies Maximize Life Expectancy?” *Public Policy & Aging Report* 31(1):7–13. doi: 10.1093/ppar/praa035.

Montez, Jennifer Karas, Mark D. Hayward, and Douglas A. Wolf. 2017. “Do U.S. States’ Socioeconomic and Policy Contexts Shape Adult Disability?” *Social Science and Medicine* 178:115–26. doi: 10.1016/j.socscimed.2017.02.012.

Montez, Jennifer Karas, Mark D. Hayward, and Anna Zajacova. 2019. “Educational Disparities in Adult Health: U.S. States as Institutional Actors on the Association.” *Socius: Sociological Research for a Dynamic World* 5:237802311983534. doi: 10.1177/2378023119835345.

Montez, Jennifer Karas, Mark D. Hayward, and Anna Zajacova. 2021. “Trends in U.S. Population Health: The Central Role of Policies, Politics, and Profits.” *Journal of Health and Social Behavior* 62(3):286–301. doi: 10.1177/00221465211015411.

Mungas, Dan, Brandon Gavett, Evan Fletcher, Sarah Tomaszewski Farias, Charles DeCarli, and Bruce Reed. 2018. “Education Amplifies Brain Atrophy Effect on Cognitive Decline: Implications for Cognitive Reserve.” *Neurobiology of Aging* 68:142–50. doi: 10.1016/j.neurobiolaging.2018.04.002.

Murman, Daniel L. 2015. “The Impact of Age on Cognition.” *Seminars in Hearing* 36(3):111–21.

Muthén, Bengt, and Tihomir Asparouhov. 2018. “Recent Methods for the Study of Measurement Invariance With Many Groups: Alignment and Random Effects.” *Sociological Methods and Research* 47(4):637–64. doi: 10.1177/0049124117701488.

Muthén, Linda K., and Bengt O. Muthén. 2017. *Mplus User’s Guide*. Eighth. Los Angeles, CA: Muthén & Muthén.

Nianogo, Roch A., Amy Rosenwohl-Mack, Kristine Yaffe, Anna Carrasco, Coles M. Hoffmann, and Deborah E. Barnes. 2022. “Risk Factors Associated with Alzheimer Disease and Related Dementias by Sex and Race and Ethnicity in the US.” Pp. 584–91 in *JAMA Neurology*. Vol. 79. American Medical Association.

Nichols, Emma, Jaimie D. Steinmetz, Stein Emil Vollset, Kai Fukutaki, Julian Chalek, Foad Abd-Allah, Amir Abdoli, Ahmed Abualhasan, Eman Abu-Gharbieh, Tayyaba Tayyaba Akram, Hanadi Al Hamad, Fares Alahdab, Fahad Mashhour Alanezi, Vahid Alipour, Sami Almustanyir, Hubert Amu, Iman Ansari, Jalal Arabloo, Tahira Ashraf, Thomas Astell-Burt, Getinet Ayano, Jose L. Ayuso-Mateos, Atif Amin Baig, Anthony Barnett, Amadou Barrow, Bernhard T. Baune, Yannick Béjot, Woldesellassie M. Mequanint Bezabhe, Yihienew Mequanint Bezabih, Akshaya Srikanth Bhagavathula, Sonu Bhaskar, Krittika Bhattacharyya, Ali Bijani, Atanu Biswas, Srinivasa Rao Bolla, Archith Boloor, Carol Brayne, Hermann Brenner, Katrin Burkart, Richard A. Burns, Luis Alberto Cámera, Chao Cao, Felix Carvalho, Luis F. S. Castro-de-Araujo, Ferrán Catalá-López, Ester Cerin, Prachi P. Chavan, Nicolas Cherbuin, Dinh Toi Chu, Vera Marisa Costa, Rosa A. S. Couto, Omid Dadras, Xiaochen Dai, Lalit Dandona, Rakhi Dandona, Vanessa De la Cruz-Góngora, Deepak Dhamnetiya, Diana Dias da Silva, Daniel Diaz, Abdel Douiri, David Edvardsson, Michael Ekholuenetale, Iman El Sayed, Shaimaa I. El-Jaafary, Khalil Eskandari, Sharareh Eskandarieh, Saman Esmaeilnejad, Jawad Fares, Andre Faro, Umar Farooque, Valery L. Feigin, Xiaoqi Feng, Seyed Mohammad Fereshtehnejad, Eduarda Fernandes, Pietro Ferrara, Irina Filip, Howard Fillit, Florian Fischer, Shilpa Gaidhane, Lucia Galluzzo, Ahmad Ghashghaee, Nermin Ghith, Alessandro Gialluisi, Syed Amir Gilani, Ionela Roxana Glavan, Elena V. Gnedovskaya, Mahaveer Golechha, Rajeev Gupta, Veer Bala Gupta, Vivek Kumar Gupta, Mohammad Rifat Haider, Brian J. Hall, Samer Hamidi, Asif Hanif, Graeme J. Hankey, Shafiul Haque, Risky Kusuma Hartono, Ahmed I. Hasaballah, M. Tasdik Hasan, Amr Hassan, Simon I. Hay, Khezar Hayat, Mohamed I. Hegazy, Golnaz Heidari, Reza Heidari-Soureshjani, Claudiu Herteliu, Mowafa Househ, Rabia Hussain, Bing Fang Hwang, Licia Iacoviello, Ivo Iavicoli, Olayinka Stephen Ilesanmi, Irena M. Ilic, Milena D. Ilic, Seyed Sina Naghibi Irvani, Hiroyasu Iso, Masao Iwagami, Roxana Jabbarinejad, Louis Jacob, Vardhmaan Jain, Sathish Kumar Jayapal, Ranil Jayawardena, Ravi Prakash Jha, Jost B. Jonas, Nitin Joseph, Rizwan Kalani, Amit Kandel, Himal Kandel, André Karch, Ayele Semachew Kasa, Gizat M. Kassie, Pedram Keshavarz, Moien AB Khan, Mahalaqua Nazli Khatib, Tawfik Ahmed Muthafer Khoja, Jagdish Khubchandani, Min Seo Kim, Yun Jin Kim, Adnan Kisa, Sezer Kisa, Mika Kivimäki, Walter J. Koroshetz, Ai Koyanagi, G. Anil Kumar, Manasi Kumar, Hassan Mehmood Lak, Matilde Leonardi, Bingyu Li, Stephen S. Lim, Xuefeng Liu, Yuewei Liu, Giancarlo Logroscino, Stefan Lorkowski, Giancarlo Lucchetti, Ricardo Lutzky Saute, Francesca Giulia Magnani, Ahmad Azam Malik, João Massano, Man Mohan Mehndiratta, Ritesh G. Menezes, Atte Meretoja, Bahram Mohajer,

Norlinah Mohamed Ibrahim, Yousef Mohammad, Arif Mohammed, Ali H. Mokdad, Stefania Mondello, Mohammad Ali Ali Moni, Md Moniruzzaman, Tilahun Belete Mossie, Gabriele Nagel, Muhammad Naveed, Vinod C. Nayak, Sandhya Neupane Kandel, Trang Huyen Nguyen, Bogdan Oancea, Nikita Otstavnov, Stanislav S. Otstavnov, Mayowa O. Owolabi, Songhomitra Panda-Jonas, Fatemeh Pashazadeh Kan, Maja Pasovic, Urvish K. Patel, Mona Pathak, Mario F. P. Peres, Arokiasamy Perianayagam, Carrie B. Peterson, Michael R. Phillips, Marina Pinheiro, Michael A. Piradov, Constance Dimity Pond, Michele H. Potashman, Faheem Hyder Pottoo, Sergio I. Prada, Amir Radfar, Alberto Raggi, Fakher Rahim, Mosiur Rahman, Pradhum Ram, Priyanga Ranasinghe, David Laith Rawaf, Salman Rawaf, Nima Rezaei, Aziz Rezapour, Stephen R. Robinson, Michele Romoli, Gholamreza Roshandel, Ramesh Sahathevan, Amirhossein Sahebkar, Mohammad Ali Sahraian, Brijesh Sathian, Davide Sattin, Monika Sawhney, Mete Saylan, Silvia Schiavolin, Allen Seylani, Feng Sha, Masood Ali Shaikh, K. S. Shaji, Mohammed Shannawaz, Jeevan K. Shetty, Mika Shigematsu, Jae Il Shin, Rahman Shiri, Diego Augusto Santos Silva, João Pedro Silva, Renata Silva, Jasvinder A. Singh, Valentin Yurievich Skryabin, Anna Aleksandrovna Skryabina, Amanda E. Smith, Sergey Soshnikov, Emma Elizabeth Spurlock, Dan J. Stein, Jing Sun, Rafael Tabarés-Seisdedos, Bhaskar Thakur, Binod Timalsina, Marcos Roberto Tovani-Palone, Bach Xuan Tran, Gebiyaw Wudie Tsegaye, Sahel Valadan Tahbaz, Pascual R. Valdez, Narayanaswamy Venketasubramanian, Vasily Vlassov, Giang Thu Vu, Linh Gia Vu, Yuan Pang Wang, Anders Wimo, Andrea Sylvia Winkler, Lalit Yadav, Seyed Hossein Yahyazadeh Jabbari, Kazumasa Yamagishi, Lin Yang, Yuichiro Yano, Naohiro Yonemoto, Chuanhua Yu, Ismaeel Yunusa, Siddhesh Zadey, Mikhail Sergeevich Zastrozhin, Anasthasia Zastrozhina, Zhi Jiang Zhang, Christopher J. L. Murray, and Theo Vos. 2022. “Estimation of the Global Prevalence of Dementia in 2019 and Forecasted Prevalence in 2050: An Analysis for the Global Burden of Disease Study 2019.” *The Lancet Public Health* 7(2):e105–25. doi: 10.1016/S2468-2667(21)00249-8.

Nuru-Jeter, Amani, Tyan Parker Dominguez, Wizdom Powell Hammond, Janxin Leu, Marilyn Skaff, Susan Egerter, Camara P. Jones, and Paula Braveman. 2009. “‘It’s the Skin You’re in’: African-American Women Talk about Their Experiences of Racism. An Exploratory Study to Develop Measures of Racism for Birth Outcome Studies.” *Maternal and Child Health Journal* 13(1):29–39. doi: 10.1007/S10995-008-0357-X/METRICS.

Oeppen, Jim, and James W. Vaupel. 2002. “Demography: Broken Limits to Life Expectancy.” *Science* 296(5570):1029–31.

Ofstedal, Mary Beth, Gwenith G. Fisher, A. Regula Herzog, Robert B. Wallace, David R. Weir, Kenneth M. Langa, Jessica D. Faul, Diane Steffick, and Stephanie Fonda. 2005. *HRS/AHEAD Documentation Report: Documentation of Cognitive Functioning Measures in the Health and Retirement Study*. Ann Arbor.

Oi, Katsuya, and Steven Haas. 2019. “Cardiometabolic Risk and Cognitive Decline: The Role of Socioeconomic Status in Childhood and Adulthood.” *Journal of Health and Social Behavior* 60(3):326–43. doi: 10.1177/0022146519867924.

Opdebeeck, Carol, Anthony Martyr, and Linda Clare. 2016. “Cognitive Reserve and Cognitive Function in Healthy Older People: A Meta-Analysis.” *Aging, Neuropsychology, and Cognition* 23(1):40–60.

Paek, Insu, Mengyao Cui, Neşe Öztürk Gübeş, and Yanyun Yang. 2018. “Estimation of an IRT Model by Mplus for Dichotomously Scored Responses Under Different Estimation Methods.” *Educational and Psychological Measurement* 78(4):569–88. doi: 10.1177/0013164417715738.

Perez, Amanda D., Suzanne M. Dufault, Erica C. Spears, David H. Chae, Cheryl L. Woods-Giscombe, and Amani M. Allen. 2023. “Superwoman Schema and John Henryism among African American Women: An Intersectional Perspective on Coping with Racism.” *Social Science and Medicine* 316. doi: 10.1016/j.socscimed.2022.115070.

Peterson, Rachel L., Kristen M. George, Lisa L. Barnes, Paola Gilsanz, Elizabeth Rose Mayeda, M. Maria Glymour, Dan M. Mungas, and Rachel A. Whitmer. 2021. “Association of Timing of School Desegregation in the United States with Late-Life Cognition in the Study of Healthy Aging in African Americans (STAR) Cohort.” *JAMA Network Open* 4(10). doi: 10.1001/jamanetworkopen.2021.29052.

Pillinger, Rebecca. 2020. “Random Slope Models.” *University of Bristol Centre for Multilevel Modeling*. Retrieved October 26, 2023 (https://www.bristol.ac.uk/cmm/learning/videos/random-slopes.html).

Pohl, Daniel J., Dominika Seblova, Justina F. Avila, Karen A. Dorsman, Erin R. Kulick, Joan A. Casey, and Jennifer Manly. 2021. “Relationship between Residential Segregation, Later-Life Cognition, and Incident Dementia across Race/Ethnicity.” *International Journal of Environmental Research and Public Health* 18(21). doi: 10.3390/ijerph182111233.

Porsteinsson, A. P., R. S. Isaacson, Sean Knox, M. N. Sabbagh, and I. Rubino. 2021. “Diagnosis of Early Alzheimer’s Disease: Clinical Practice in 2021.” *Journal of Prevention of Alzheimer’s Disease* 8(3):371–86.

Power, Melinda C., Erin E. Bennett, Robert W. Turner, N. Maritza Dowling, Adam Ciarleglio, M. Maria Glymour, and Kan Z. Gianattasio. 2021. “Trends in Relative Incidence and Prevalence of Dementia across Non-Hispanic Black and White Individuals in the United States, 2000-2016.” *JAMA Neurology* 78(3):275–84. doi: 10.1001/jamaneurol.2020.4471.

Prince, Martin, Anders Wimo, Maëlenn Guerchet, Gemma-Claire Ali, Yu-Tzu Wu, Matthew Prina, and Alzheimer’s Disease International. 2015. *World Alzheimer Report 2015: The Global Impact of Dementia, An Analysis of Prevalence, Incidence, Cost and Trends*. London.

Qiu, Chengxuan, and Laura Fratiglioni. 2018. “Aging without Dementia Is Achievable: Current Evidence from Epidemiological Research.” *Journal of Alzheimer’s Disease* 62(3):933–42.

van Raalte, Alyson A., and Hal Caswell. 2013. “Perturbation Analysis of Indices of Lifespan Variability.” *Demography* 50(5):1615–40. doi: 10.1007/s13524-013-0223-3.

van Raalte, Alyson A., and Pekka Martikainen. 2014. “Lifespan Variation by Occupational Class: Compression or Stagnation Over Time?” *Demography* 51(1):73–95.

van Raalte, Alyson, Isaac Sasson, and Pekka Martikainen. 2018. “The Case for Monitoring Lifespan Inequality.” *Science* 362(6418):1002–4.

Rajan, Kumar B., Jennifer Weuve, Lisa L. Barnes, Elizabeth A. McAninch, Robert S. Wilson, and Denis A. Evans. 2021. “Population Estimate of People with Clinical Alzheimer’s

Disease and Mild Cognitive Impairment in the United States (2020–2060)." *Alzheimer's and Dementia* 17(12):1966–75. doi: 10.1002/alz.12362.

Rajan, Kumar B., Jennifer Weuve, Lisa L. Barnes, Robert S. Wilson, and Denis A. Evans. 2019. "Prevalence and Incidence of Clinically Diagnosed Alzheimer's Disease Dementia from 1994 to 2012 in a Population Study." *Alzheimer's and Dementia* 15(1):1–7. doi: 10.1016/j.jalz.2018.07.216.

RAND Corporation. 2022. "RAND HRS Longitudinal File 2018 (V2)."

Rigdon, Edward E. 1996. "CFI versus RMSEA: A Comparison of Two Fit Indexes for Structural Equation Modeling." *Structural Equation Modeling* 3(4):369–79. doi: 10.1080/10705519609540052.

Robert Wood Johnson Foundation. 2017. *Discrimination in America: Experiences and Views of African Americans*.

Rodriguez, Javier M., Arun S. Karlamangla, Tara L. Gruenewald, Dana Miller-Martinez, Sharon S. Merkin, and Teresa E. Seeman. 2019. "Social Stratification and Allostatic Load: Shapes of Health Differences in the MIDUS Study in the United States." *Journal of Biosocial Science* 51(5):627–44. doi: 10.1017/S0021932018000378.

Roehr, Susanne, Alexander Pabst, Tobias Luck, and Steffi G. Riedel-Heller. 2018. "Is Dementia Incidence Declining in High-Income Countries? A Systematic Review and Meta-Analysis." *Clinical Epidemiology* 10:1233–47.

Ross, Lori E., Laurel O'Gorman, Melissa A. MacLeod, Greta R. Bauer, Jenna MacKay, and Margaret Robinson. 2016. "Bisexuality, Poverty and Mental Health: A Mixed Methods Analysis." *Social Science and Medicine* 156:64–72. doi: 10.1016/j.socscimed.2016.03.009.

Salthouse, Timothy A. 2009. "When Does Age-Related Cognitive Decline Begin?" *Neurobiology of Aging* 30(4):507–14.

Salthouse, Timothy A. 2010. "Selective Review of Cognitive Aging." *Journal of the International Neuropsychological Society* 16(5):754–60.

Salthouse, Timothy A. 2019. "Trajectories of Normal Cognitive Aging." *Psychology and Aging* 34(1):17–24. doi: 10.1037/pag0000288.

Sass, Daniel A. 2011. "Testing Measurement Invariance and Comparing Latent Factor Means within a Confirmatory Factor Analysis Framework." *Journal of Psychoeducational Assessment* 29(4):347–63. doi: 10.1177/0734282911406661.

Sasson, Isaac. 2016. "Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010." *Demography* 53(2):269–93. doi: 10.1007/s13524-015-0453-7.

Schmitz, Norbert, Sonya S. Deschênes, Rachel J. Burns, Sofia M. Danna, Oscar H. Franco, M. Arfan Ikram, Mika Kivimäki, Archana Singh-Manoux, and Henning Tiemeier. 2018. "Cardiometabolic Dysregulation and Cognitive Decline: Potential Role of Depressive Symptoms." *The British Journal of Psychiatry* 212(2):96–102. doi: 10.1192/bjp.2017.26.

Schneider, Edward L., and Jacob A. Brody. 1983. "Aging, Natural Death, and the Compression of Morbidity: Another View." *The New England Journal of Medicine* 309(14):854–56.

Seblova, D., R. Berggren, and M. Lövdén. 2020a. "Education and Age-Related Decline in Cognitive Performance: Systematic Review and Meta-Analysis of Longitudinal Cohort Studies." *Ageing Research Reviews* 58.

Seblova, D., R. Berggren, and M. Lövdén. 2020b. "Education and Age-Related Decline in Cognitive Performance: Systematic Review and Meta-Analysis of Longitudinal Cohort Studies." *Ageing Research Reviews* 58.

Seplaki, Christopher L., Noreen Goldman, Maxine Weinstein, and Yu-Hsuan Lin. 2006. "Measurement of Cumulative Physiological Dysregulation in an Older Population." *Demography* 43(1):165–83.

Sisco, Shannon, Alden L. Gross, Regina A. Shih, Bonnie C. Sachs, M. Maria Glymour, Katherine J. Bangen, Andreana Benitez, Jeannine Skinner, Brooke C. Schneider, and Jennifer J. Manly. 2015. "The Role of Early-Life Educational Quality and Literacy in Explaining Racial Disparities in Cognition in Late Life." *Journals of Gerontology - Series B Psychological Sciences and Social Sciences* 70(4):557–67. doi: 10.1093/geronb/gbt133.

Sonnega, Amanda, Jessica D. Faul, Mary Beth Ofstedal, Kenneth M. Langa, John W. R. Phillips, and David R. Weir. 2014. "Cohort Profile: The Health and Retirement Study (HRS)." *International Journal of Epidemiology* 43(2):576–85. doi: 10.1093/ije/dyu067.

Stata Corp. 2021. "Stata Statistical Software: Release 17."

Steenland, Kyle, Felicia C. Goldstein, Allan Levey, and Whitney Wharton. 2016. "A Meta-Analysis of Alzheimer's Disease Incidence and Prevalence Comparing African-Americans and Caucasians." *Journal of Alzheimer's Disease* 50(1):71–76. doi: 10.3233/JAD-150778.

Steptoe, Andrew, Mark Hamer, and Yoichi Chida. 2007. "The Effects of Acute Psychological Stress on Circulating Inflammatory Factors in Humans: A Review and Meta-Analysis." *Brain Behav Immun* 21(7):901–12. doi: https://doi.org/10.1016/j.bbi.2007.03.011.

Stern, Yaakov. 2002. "What Is Cognitive Reserve? Theory and Research Application of the Reserve Concept." *Journal of the International Neuropsychological Society* 8(3):448–60. doi: 10.1017/S1355617702813248.

Stern, Yaakov. 2012. *Cognitive Reserve in Ageing and Alzheimer's Disease*. Vol. 11.

Stern, Yaakov, Eider M. Arenaza-Urquijo, David Bartrés-Faz, Sylvie Belleville, Marc Cantilon, Gael Chetelat, Michael Ewers, Nicolai Franzmeier, Gerd Kempermann, William S. Kremen, Ozioma Okonkwo, Nikolaos Scarmeas, Anja Soldan, Chinedu Udeh-Momoh, Michael Valenzuela, Prashanthi Vemuri, Eero Vuoksimaa, Eider M. Arenaza Urquiljo, Marc Cantillon, Sean A. P. Clouston, Ainara Estanga, Brian Gold, Christian Habeck, Richard Jones, Renata Kochhann, Yen Ying Lim, Pablo Martínez-Lage, Silvia Morbelli, Ozioma Okonkwo, Rik Ossenkoppele, Corinne Pettigrew, Allyson C. Rosen, Xiaowei Song, and Anita C. Van Loenhoud. 2020. "Whitepaper: Defining and Investigating Cognitive Reserve, Brain Reserve, and Brain Maintenance." *Alzheimer's and Dementia* 16(9):1305–11.

Sternthal, Michelle J., Natalie Slopen, and David R. Williams. 2011. "Racial Disparities in Health: How Much Does Stress Really Matter?" *Du Bois Review: Social Science Research on Race* 8(1):95–113. doi: 10.1017/S1742058X11000087.

Sue, Derald Wing, Christina M. Capodilupo, Gina C. Torino, Jennifer M. Bucceri, Aisha M. B. Holder, Kevin L. Nadal, and Marta Esquilin. 2007. "Racial Microaggressions in Everyday Life: Implications for Clinical Practice." *American Psychologist* 62(4):271–86. doi: 10.1037/0003-066X.62.4.271.

Svetina, Dubravka, Leslie Rutkowski, and David Rutkowski. 2020. "Multiple-Group Invariance with Categorical Outcomes Using Updated Guidelines: An Illustration Using Mplus and the Lavaan/SemTools Packages." *Structural Equation Modeling* 27(1):111–30. doi: 10.1080/10705511.2019.1602776.

Tamir, Christine. 2021. *The Growing Diversity of Black America*. Washington, DC.

Thomas, Anita Jones, Karen M. Witherspoon, and Suzette L. Speight. 2008. "Gendered Racism, Psychological Distress, and Coping Styles of African American Women." *Cultural Diversity and Ethnic Minority Psychology* 14(4):307–14. doi: 10.1037/1099-9809.14.4.307.

Tolnay, Stewart E. 2003. "The African American 'Great Migration' and Beyond." *Annual Review of Sociology* 29:209–32.

Topping, Michael, Jinho Kim, and Jason Fletcher. 2021. "Geographic Variation in Alzheimer's Disease Mortality." *PLoS ONE* 16(7 July). doi: 10.1371/journal.pone.0254174.

Tucker-Drob, Elliot M. 2019. "Cognitive Aging and Dementia: A Life-Span Perspective." *Annual Review of Developmental Psychology* 1:177–96. doi: 10.1146/annurev-devpsych-121318.

Tuljapurkar, Shripad. 2011. "The Final Inequality: Variance in Age at Death." Pp. 209–21 in *Demography and the Economy*, edited by J. B. Shoven. Chicago, IL: University of Chicago Press.

U.S. Census Bureau. 2018. "Race and Hispanic Origin by Selected Age Groups: Main Projections Series for the United States, 2017-2060."

U.S. Census Bureau. 2021. "Race and Ethnicity in the United States: 2010 Census and 2020 Census." Retrieved March 22, 2023 (https://www.census.gov/library/visualizations/interactive/race-and-ethnicity-in-the-united-state-2010-and-2020-census.html).

Vespa, Jonathan, Lauren Medina, and David M. Armstrong. 2020. *Demographic Turning Points for the United States: Population Projections for 2020 to 2060*. Washington, DC.

Wang, Jichuan, and Xiaoqian Wang. 2012. *Structural Equation Modeling: Applications Using Mplus*. 1st ed. John Wiley & Sons.

Weuve, Jennifer, Lisa L. Barnes, Carlos F. Mendes De Leon, Kumar B. Rajan, Todd Beck, Neelum T. Aggarwal, Liesi E. Hebert, David A. Bennett, Robert S. Wilson, and Denis A. Evans. 2018. "Cognitive Aging in Black and White Americans: Cognition, Cognitive Decline, and Incidence of Alzheimer Disease Dementia." *Epidemiology* 29(1):151–59. doi: 10.1097/EDE.0000000000000747.

Wicherts, Jelte M. 2016. "The Importance of Measurement Invariance in Neurocognitive Ability Testing." *Clinical Neuropsychologist* 30(7):1006–16.

Wicherts, Jelte M., and Conor V. Dolan. 2010. “Measurement Invariance in Confirmatory Factor Analysis: An Illustration Using IQ Test Performance of Minorities.” *Educational Measurement: Issues and Practice* 29(3):39–47. doi: 10.1111/j.1745-3992.2010.00182.x.

Widaman, Keith F., and Steven P. Reise. 1997. “Exploring the Measurement Invariance of Psychological Instruments: Applications in the Substance Use Domain.” Pp. 281–324 in *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research*, edited by K. J. Bryant, M. T. Windle, and S. G. West. American Psychological Association.

Williams, David R. 1999. “Race, Socioeconomic Status, and Health the Added Effects of Racism and Discrimination.” *Annals of the New York Academy of Sciences* 896:173–88. doi: 10.1111/j.1749-6632.1999.tb08114.x.

Williams, David R. 2012. “Miles to Go before We Sleep: Racial Inequities in Health.” *J Health Soc Behav* 53(3):279–95. doi: 10.1177/0022146512455804.

Williams, David R., and Chiquita Collins. 2001. “Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health.” *Public Health Reports* 116(5):404–16. doi: 10.1093/phr/116.5.404.

Williams, David R., and Selina A. Mohammed. 2013. “Racism and Health I: Pathways and Scientific Evidence.” *American Behavioral Scientist* 57(8):1152–73. doi: 10.1177/0002764213487340.

Williams, David R., Naomi Priest, and Norman B. Anderson. 2016. “Understanding Associations among Race, Socioeconomic Status, and Health: Patterns and Prospects.” *Health Psychology* 35(4):407–11. doi: 10.1037/hea0000242.

Williams, David R., and Michelle Sternthal. 2010. “Understanding Racial-Ethnic Disparities in Health: Sociological Contributions.” *J Health Soc Behav* 51(1_suppl):S15–27.

Wilson, Christopher J., Stephen C. Bowden, Linda K. Byrne, Nicole R. Joshua, Wolfgang Marx, and Lawrence G. Weiss. 2023. “The Cross-Cultural Generalizability of Cognitive Ability Measures: A Systematic Literature Review.” *Intelligence* 98.

Woods-Giscombé, Cheryl L. 2010. “Superwoman Schema: African American Womens Views on Stress, Strength, and Health.” *Qualitative Health Research* 20(5):668–83. doi: 10.1177/1049732310361892.

World Health Organization. 2011. *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus* . Geneva, Switzerland.

Yang, Yang C., Christine E. Walsh, Kaitlin Shartle, Rebecca C. Stebbins, Allison E. Aiello, Daniel W. Belsky, Kathleen Mullan Harris, Marianne Chanti-Ketterl, and Brenda L. Plassman. 2023. “An Early and Unequal Decline: Life Course Trajectories of Cognitive Aging in the United States.” *Journal of Aging and Health*. doi: 10.1177/08982643231184593.

Zacher, Meghan, Samantha Brady, and Susan E. Short. 2023. “Geographic Patterns of Dementia in the United States: Variation by Place of Residence, Place of Birth, and Subpopulation.” Pp. 1192–1203 in *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*. Vol. 78. Gerontological Society of America.

Zuelsdorff, Megan, Ozioma C. Okonkwo, Derek Norton, Lisa L. Barnes, Karen L. Graham, Lindsay R. Clark, Mary F. Wyman, Susan F. Benton, Alexander Gee, Nickolas Lambrou, Sterling C. Johnson, Carey E. Gleason, and Laura Zahodne. 2020. "Stressful Life Events and Racial Disparities in Cognition among Middle-Aged and Older Adults." *Journal of Alzheimer's Disease* 73(2):671–82. doi: 10.3233/JAD-190439.

Zwick, Rebecca. 2019. "Assessment in American Higher Education: The Role of Admissions Tests." *Annals of the American Academy of Political and Social Science* 683(1):130–48. doi: 10.1177/0002716219843469.