Assessing bias in cognitive testing for older adults with sensory impairment: an analysis of differential item functioning in the Baltimore Longitudinal Study on Aging (BLSA) and the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS)

E. Nichols[1], **J.A. Deal**[1,2], **B.K. Swenor**[1,3], **A.G. Abraham**[1,4], **N.M. Armstrong**[5], **M.C. Carlson**[6], **M. Griswold**[7], **F.R. Lin**[1,2,6], **T.H. Mosley**[7], **P.Y. Ramulu**[3], **N.S. Reed**[1,2], **S.M. Resnick**[8], **A.R. Sharrett**[1], **A.L. Gross**[1]

[1.]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, W6508, Baltimore, MD 21205, USA

[2.]Cochlear Center for Hearing and Public Health, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA

[3.]Wilmer Eye Institute, Johns Hopkins Hospital, 1800 Orleans St, Baltimore, MD 21287, USA

[4.]Department of Epidemiology, School of Public Health, University of Colorado Anschutz Medical Campus, 1635 Aurora Ct, Aurora, CO 80045, USA

[5.]Department of Psychiatry and Human Behavior, Brown University Warren Alpert Medical School, 700 Butler Dr, Box G-BH, Providence, RI 02906, USA

[6.]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway St, 8th Floor, Baltimore, MD 21205, USA

[7.]Memory Impairment and Neurodegenerative Dementia Center, University of Mississippi Medical Center, 2500 North State St, Jackson, MS 39216, USA

[8.]Laboratory of Behavioral Neuroscience, National Institute on Aging, 251 Bayview Blvd, Suite 101, Baltimore, MD 21224, USA

Abstract

Objective: Vision and hearing impairments affect 55% of people aged 60+ and are associated with lower cognitive test performance; however tests rely on vision, hearing, or both. We hypothesized scores on tests that depend on vision or hearing are different among those with vision or hearing impairments, respectively, controlling for underlying cognition.

Corresponding Author: Emma Nichols, 2024 E Monument St Suite 2-700, Tel: 410-955-0491, Fax: 410-614-9625, enicho30@jh.edu.

Method: Leveraging cross-sectional data from the Baltimore Longitudinal Study of Aging (BLSA) and the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS), we used item response theory to test for differential item functioning (DIF) by vision impairment (better-eye presenting visual acuity worse than 20/40) and hearing impairment (better-ear 0.5-4 kHz pure-tone average>25 decibels).

Results: We identified DIF by vision impairment for tests whose administrations do not rely on vision (e.g., Delayed Word Recall in both ARIC-NCS: 0.50 logit difference between impaired and unimpaired (p=0.04) and in BLSA: 0.62 logits (p=0.02)) and DIF by hearing impairment for tests whose administrations do not rely on hearing (Digit Symbol Substitution test in BLSA: 1.25 logits (p=0.001) and Incidental Learning test in ARIC-NCS: 0.35 logits (p=0.001)). However, no individuals had differences between unadjusted and DIF-adjusted measures of greater than the standard error of measurement.

Conclusions: That DIF by sensory impairment in cognitive tests is independent of administration characteristics could indicate elevated cognitive load among persons with sensory impairment plays a larger role in test performance than previously acknowledged. While these results were unexpected, neither of these samples are nationally representative and each has unique selection factors, thus replication is critical.

Introduction

Recent estimates of the prevalence of Alzheimer's disease in the United States range from 2.8-5.8 million (Alzheimer's Association, 2019; Brookmeyer, Johnson, Ziegler-Graham, & Arrighi, 2007; Hebert, Weuve, Scherr, & Evans, 2013; Plassman et al., 2007). These estimates and their corresponding uncertainty are affected by potential methodological variations, including inconsistencies in sampling methods and case definitions. One important source of variation may stem from how different subgroups respond to tests in neuropsychological batteries.

The questions or items used in these batteries rely to varying degrees on visual stimuli, auditory stimuli or both. Approximately 14 million Americans have visual impairment, while an estimated 22 million Americans 60 years have a clinically significant hearing impairment (Goman & Lin, 2016; Frank R. Lin, Niparko, & Ferrucci, 2011; Vitale, Cotch, & Sperduto, 2006). Those with either impairment comprise 55% of Americans 60 years overall (Swenor, Ramulu, Willis, Friedman, & Lin, 2013). Hence, the potential effect of biases in cognitive testing that relies on either vision or hearing for the modality of administration could be substantial.

An understanding of this potential bias is particularly important, given the large body of cross-sectional evidence supporting the existence of associations between sensory impairments and lower cognitive test scores (Chen, Bhattacharya, & Pershing, 2017; Gussekloo, de Craen, Oduber, van Boxtel, & Westendorp, 2005; Taljaard, Olaithe, Brennan-

Jones, Eikelboom, & Bucks, 2016). Additionally, there has been growing evidence from longitudinal studies suggesting strong associations of vision loss (M. Y. Lin et al., 2004; Rogers & Langa, 2010; Tay et al., 2006; Zheng et al., 2018) and hearing (Deal et al., 2017; Frank R. Lin, Ferrucci, et al., 2011) with accelerated cognitive decline and dementia. Proposed explanations for these associations include the possibility of common pathways, i.e., microvascular disease, but some evidence also supports causal pathways via mechanisms, i.e., social isolation or changes in brain structure and function (F. R. Lin et al., 2014; Frank R. Lin & Albert, 2014; H. Lin et al., 2018; Peelle, Troiani, Grossman, & Wingfield, 2011). Support for these proposed causal mechanisms would suggest that interventions targeted at correcting vision loss and hearing impairment could have important implications for preventing cognitive decline and dementia (Ehrlich & Langa, 2019; Frank R. Lin & Albert, 2014).

Prior literature comparing standard cognitive test battery scores with scores removing items dependent on adequate vision or hearing does suggest that hearing impairment in particular may bias cognitive testing (Dupuis et al., 2015). However, this work utilized a brief battery (the Montreal Cognitive Assessment) and relied on the assumption that the further abbreviated batteries removing items dependent on hearing or vision still adequately described global cognitive performance. Additionally, they were unable to evaluate bias by individual cognitive test items. Overall, it remains largely unknown what effect biases in cognitive instruments might have on our understanding of the relationship between sensory impairment and cognition.

Item response theory methods are commonly used to characterize the measurement of traits, such as cognitive function, that are not directly observable (Bollen, 1989). These methods have been previously applied to investigate how language of administration and demographic differences, such as sex or educational level, may bias cognitive test results, while accounting for underlying cognitive ability (Jones, 2006; Jones & Gallo, 2002; Morales, Flowers, Gutierrez, Kleinman, & Teresi, 2006). However, these methods have not been heretofore applied to the study of sensory impairment. This study aims to apply these methods to better understand the effect of sensory impairment on the measurement of specific cognitive functions. We leveraged data from two well-characterized, prospective cohort studies, the Baltimore Longitudinal Study of Aging (BLSA) and the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). We hypothesized that cognitive tests where administration relies on vision (e.g., Trail Making test, parts A & B), or hearing (e.g., auditory word-list learning) would be more difficult (biased) for those with vision and hearing impairments, respectively, even if they did not have reduced cognitive ability.

Methods

Study Populations

BLSA is a prospective longitudinal study of psychological and physical aging that was initiated in 1958. It recruited healthy, community-dwelling volunteers to study aging over the adult lifespan. Participants undergo cognitive, psychological and physical testing at the National Institute on Aging, in Baltimore, Maryland (Shock et al., 1984). Our study utilized data from participants 55 years and older who attended visits between 2015 and 2017, who

had cognitive testing and either objective vision or hearing testing, as described below. Analyses of cognition by vision status used data from individuals who had measurements of cognition and vision (N=703). Analyses of cognition by hearing status used individuals who had measurements of cognition and hearing (N=671).

The ARIC-NCS study recruited individuals at four university-based study sites (Forsyth County, North Carolina; Jackson, Mississippi; the northwest suburbs of Minneapolis, Minnesota; and Washington County, Maryland) aged 45 to 64 years from 1987 to 1989 and followed them for future health outcomes (The ARIC investigators, 1989). Vision (Jackson and Washington county sites as part of the Eye Determinants of Cognition [EyeDOC] ancillary study) and hearing (all sites) assessments took place from 2017 to 2019 and our analytic sample included those who had data on cognition and vision (N = 1,051) or data on cognition and hearing (N = 3,591).

All participants provided written informed consent and the study protocols were approved by either the National Institute of Environmental Health Sciences Review Board (BLSA) or the site specific Institutional Review Boards at each study center (ARIC).

All analyses were conducted separately in the BLSA and ARIC-NCS samples and results were compared across the samples to examine similarities and differences in the patterns of findings.

Measures of Sensory Impairment

Vision was assessed monocularly using an ETDRS (Early Treatment of Diabetic Retinopathy) chart following standardized protocols (Ferris III, Kassoff, Bresnick, & Bailey, 1982) in BLSA and the EyeDOC sub-study of the ARIC-NCS. Participants wore their normal corrective lenses (if any) and the number of letters read was recorded and used to calculate better-eye distance visual acuity (in logMAR). Ambient light was measured and a standard lightbox was used to ensure valid testing could be conducted. For this study, visual impairment was defined as presenting better-eye visual acuity worse than 20/40, following guidance of the American Academy of Ophthalmology (AAO).

Pure-tone audiometry was used to measure hearing in both BLSA and ARIC-NCS. This is the gold-standard method for evaluating the faintest detectable tones for a range of pitches. The measure is not reliant on higher-order cognition (Pickles, 2012). All participants in BLSA and ARIC-NCS were tested in a sound-attenuating booth that met American National Standards Institute (ANSI) standards and pure-tone conduction thresholds were obtained using an Interacoustics AD629 or Equinox audiometer (Interacoustics A/S, Assens, Denmark). For those participants assessed during home visits in ARIC-NCS, pure-tone audiometry was conducted with a portable audiometer (Shoebox audiometry, Ottawa, Canada). Ambient level of noise was assessed to ensure that valid testing could be conducted. We defined hearing impairment as having a 4-frequency (0.5, 1, 2, 4 kHz) pure tone average of 25 decibels hearing level or higher in the better-hearing ear ("Prevention of Blindness and Deafness Grades of Hearing Impairment," n.d.).

Neuropsychological Assessments

Cognitive tests across BLSA and ARIC-NCS cover a range of cognitive domains, and rely to varying degrees on vision, hearing, or both. Attentional ability was captured by the Trail-Making test, part A in both BLSA and ARIC-NCS and additionally by the Weschler Adult Intelligence Scale– Revised (WAIS-R) digits forward task in BLSA (Blackburn & Benton, 1957; R. M. Reitan, 1986). BLSA and ARIC-NCS both assessed language using semantic and phonemic fluency, as well as the Boston Naming Test (Benton & Hamsher, 1976; Williams, Mack, & Henderson, 1989). Delayed free recall was assessed in both BLSA (California Verbal Learning Test [CVLT] long delay) and ARIC-NCS (Delayed Recall Task) to assess verbal memory, and BLSA additionally included tests of immediate free recall for words (CVLT short delay) and for visually presented figures (Benton Visual Retention Test) (Benton, 1962; Delis, Kramer, Kaplan, & Thompkins, 1987). We excluded the test of immediate verbal memory from the BLSA analysis due to collinearity with the delayed free recall task. ARIC-NCS additionally included a test of logical memory (Wechsler Memory Scale – Revised paragraph recall test) and a test of incidental learning from the digit symbol substitution task (Smith, 1968; Wechsler, 1987). Executive function was measured using the WAIS-R digits backwards task, the digit symbol substitution task and the Trail Making test, part B in both BLSA and ARIC-NCS (Blackburn & Benton, 1957; Ralph M. Reitan, 1958; Ryan & Lopez, 2001). BLSA also included the WAIS-R Similarities task. Finally, the Card Rotation Test and a clock drawing measure were used in BLSA as additional measures of executive functioning and visuospatial ability (Table 1). Based on the features of administration in the BLSA and ARIC-NCS studies, Boston Naming Test, Trail Making Test, Parts A and B, the digit symbol substitution test, and tests of figural memory, incidental learning, card rotation and clock drawing relied primarily on vision. Administrations for category and letter fluency tests, delayed free recall in BLSA, logical memory, WAIS-R digits forward and backwards tests and the similarities task are more strongly contingent on adequate hearing (Table 1). We acknowledge that the extent to which a cognitive task relies on a sensory ability is a continuous and not a binary notion, thus these classifications are not intended to be consequential.

Statistical Analysis

We first computed descriptive statistics to characterize the demographics of the BLSA and ARIC-NCS samples. We then used item response theory methods to characterize cognition in the BLSA and ARIC-NCS samples (Lord, 1980). To prepare the data, we discretized cognitive test variables into up to 8 categories using an equal interval discretization method (Rucker, McShane, & Preacher, 2015). We collapsed adjoining categories if one category had less than 5 percent of the records in the dataset or if any indicator had 0 records in either of the groups with vision or hearing impairment (eTables 1 & 2).

Item response theory methods for testing differential item functioning (DIF) seek to disentangle the true effect of sensory impairment on cognitive function from the effect that sensory impairment may have on cognitive test performance (bias). By controlling for underlying levels of cognitive ability, defined based on the collective responses to all test items in a cognitive battery, item response theory methods facilitate the quantification of bias in test performance by sensory impairment status. We used the alignment method in

Mplus software (version 8, Muthén & Muthén, Los Angeles CA) (Asparouhov & Muthén, 2014; Muthén & Asparouhov, 2014), which allows for testing of differences between groups on each parameter estimated in the model. The elemental measurement model includes parameters describing (1) the difficulty of each categorical response to a given item (item thresholds), and (2) the strength of the association between a given item and underlying cognitive performance (item discrimination). Alignment analysis expands on this framework by estimating this elemental model in groups defined by sensory impairment status, and testing the equivalence of each of the aforementioned parameters across groups. We tested for DIF in the BLSA and ARIC-NCS samples separately. Additional details on these models are available in the Appendix.

To avoid focusing on potentially spurious results and to control for type 1 error, when inspecting DIF by threshold, we concentrated on significant findings at the ends of the threshold ranges, or in instances where multiple sequential thresholds showed significant DIF. Additionally, we examined item characteristic curves (ICCs) by hearing and vision impairment to visually describe the results of the alignment analysis.

We then calculated a measure of salient DIF to determine the impact of DIF by hearing and vision impairment on the estimation of cognition. To quantify the impact of DIF on individual-level estimates of cognitive functioning in BLSA and ARIC–NCS separately, we compared factor scores from a unidimensional confirmatory factor analysis with those from a model that allowed parameter estimates where we identified DIF to vary between either the vision impaired and normal vision groups, or the hearing impaired and normal hearing groups, thus accounting for the DIF that we observed. The difference between estimates of latent cognitive functioning in the first set of models and the second set of models quantifies absolute magnitude of the effect of DIF on estimates of cognition. We evaluated the proportion of people whose differences fall outside one standard error of measurement, as previous work has used this threshold to describe meaningful differences in the classification of cognitive impairment for clinical and research purposes (Cella, Eton, Lai, Peterman, & Merkel, 2002; Eton et al., 2004; Yost et al., 2005).

Sensitivity Analyses

Item response theory methods for the detection of DIF control for latent cognitive ability, and therefore also control for differences in demographic factors (such as race or sex) that result in differences in latent cognitive ability. However, it is possible that demographic factors may influence test performance above and beyond their influence on latent cognitive ability, which might affect our findings. To evaluate the sensitivity of our results to the effect of differences in demographic characteristics, we conducted stratified analyses in the ARIC-NCS sample due to its larger sample size. We conducted our primary DIF analysis on subsets of the ARIC-NCS sample stratified into the following groups: those above 77 years old, below 77 years old, those with low educational attainment (high school or lower), high educational attainment, white race, black race, males and females.

Results

Sample Characteristics

The BLSA sample had a mean age 74.7 years (Standard deviation, SD = 9.0) and was 34.0% black and 55.6% female. In BLSA, 25.2% of participants had a vision impairment and 52.2% of participants had a hearing impairment. The ARIC-NCS sample was older on average (mean age of 79.8 years; SD = 4.7) and was 24.0% black and 59.8% female; 18.2% had a vision impairment and 73.0% had a hearing impairment (Table 2). Educational attainment was also higher in BLSA (4.6% with high school or equivalent education, 95.1% with greater than high school) than in ARIC-NCS (41.3% with high school or equivalent education, 46.6% with greater than high school). The vision sample in ARIC-NCS, which was comprised of participants from two of the four ARIC-NCS sites, had a higher proportion of black participants (46.9%) and a higher proportion of female participants (63.2%) as compared to the overall ARIC-NCS sample who had any sensory impairment measured. There were no other substantial differences between the overall and impairment-specific samples (eTables 3 & 4). In general, people with sensory impairment tended to perform worse on most of the tests in the battery (eTables 5 & 6).

DIF by vision impairment

Table 3 presents results in BLSA and separately in ARIC-NCS for DIF by vision impairment, controlling for underlying cognitive performance. In the BLSA sample, there were two instances of DIF by vision impairment. The delayed word recall item was more difficult for those with vision impairment among individuals with low cognitive functioning (Parameter difference (PD) = 0.62, Standard error (SE) = 0.28, Standardized effect size (SES) = 2.21). Additionally, the Boston Naming Test was easier for individuals with vision impairment among those with low cognitive functioning (PD = −0.74, SE = 0.35, SES = −2.11). These findings are also displayed graphically in Figure 1 and eFigures 1-4.

When evaluating DIF by vision impairment in ARIC-NCS, we saw that delayed word recall was also more difficult in those with vision impairment, but for those at the highest thresholds of cognitive performance (PD = 0.50; SE = 0.24; SES = 2.08). Boston Naming Test was again easier for people with vision impairment, but among those at the highest levels of underlying cognitive performance (PD = −0.55; SE = 0.24; SES= −2.29). In this sample, we also observed DIF on the discrimination parameter for the digit symbol substitution task, indicating that the item's correlation with latent cognition was weaker among those with vision impairment (PD = −0.93; SE = 0.42; SES = −2.21).

DIF by hearing impairment

With respect to DIF by hearing impairment using the alignment method, we identified DIF in both the BLSA and ARIC-NCS samples, controlling for underlying cognition. In BLSA, the middle three thresholds of five for the category fluency test (animal naming) (PD = 0.75, 0.72, 0.87; SE = 0.32, 0.32, 0.39; SES = 2.34, 2.25, 2.23) and the first four thresholds of five for the digit symbol substitution test showed DIF (PD = 1.25, 0.91, 0.98, 1.01; SE = 0.37, 0.37, 0.41, 0.46; SES = 3.38, 2.46, 2.39, 2.20) (Table 4). The parameter differences (group with hearing impairment – group without hearing impaired) were positive, indicating

that this item was more difficult for those with hearing impairment, after controlling for underlying cognitive function.

In ARIC-NCS, we also saw evidence that the category fluency test (animal naming) was more difficult in participants with hearing impairment at the lowest two thresholds (PD = 0.85, 0.34, SE = 0.23, 0.14, SES = 3.70, 2.43). This indicates that this test was more difficult for participants with hearing impairment, specifically at the lower ends of the distribution of the latent cognitive trait (in those with lower cognitive functioning). The phonemic fluency task (naming words that start with F, A and S) was again more difficult in those with hearing impairment at the lowest levels of the cognitive trait distribution (PD = 0.37; SE = 0.17; SES = 2.18). The delayed word recall (PD = 0.26, 0.40, 0.38, 0.33; SE = 0.12, 0.09, 0.09, 0.10; SES = 2.17, 4.44, 4.22, 3.30) and incidental learning tasks (PD = 0.35, 0.31, 0.33; SE = 0.10, 0.10, 0.14; SES= 3.50, 3.10, 2.36), both of which measure memory, were more difficult for those with hearing impairment across a large range of thresholds. These two items also showed DIF in the discrimination parameter (delayed word recall: PD = 0.20, SE = 0.08, SES = 2.5; incidental learning: PD = 0.35, SE = 0.11, SES = 3.18), indicating that the tests were more strongly related to the underlying cognitive latent trait in the group with hearing impairment. Finally, we also saw DIF in the Boston Naming Test, such that the item was easier for those with hearing impairment at the highest levels of cognitive functioning (PD = −0.28, −0.36; SE = 0.12, 0.12; SES = −2.33, −3.00).

Sensitivity Analyses

Our sensitivity analyses revealed that our findings were fairly consistent when comparing results stratified by different demographic subgroups (eTables 7 & 8).

Salient DIF

Despite finding evidence of DIF by vision and hearing impairment status in both BLSA and ARIC-NCS, our analysis of salient DIF based on the difference between DIF-adjusted and DIF-unadjusted cognitive scores indicated that there were no instances of DIF which resulted in bias of greater than 1 standard error of measurement in estimated cognitive scores in participants with and without vision and hearing impairment respectively (Figure 2).

Discussion

This study found evidence of DIF by vision impairment status in cognitive test items that did not rely on vision as well as evidence of DIF by hearing impairment status in cognitive test items that did not rely primarily on hearing. This indicates that vision and hearing impairments may affect performance on cognitive testing, controlling for underlying cognitive ability. However, our evaluation of salient DIF indicated that these differences may not lead to substantial differences in measurement of cognitive functioning in older adults by sensory impairment status.

We were able to replicate the main pattern of findings in our results across the two independent samples (ARIC-NCS and BLSA) that we included. However, we found the most evidence of DIF by hearing and vision impairment in more cognitively impaired and less healthy subgroups in both the BLSA and ARIC-NCS samples. Our analysis of

comparative cognitive performance in BLSA and ARIC-NCS (eMethods) indicated that underlying cognitive ability was higher in BLSA as compared to the ARIC-NCS sample (eFigure 5). BLSA represents a group of volunteers, who were screened to be healthy at baseline, whereas ARIC-NCS was initially sampled to be population representative. Therefore participants in ARIC-NCS are on average less healthy, and have a larger number of comorbidities (Shock et al., 1984; The ARIC investigators, 1989). In the analysis of hearing impairment, we found evidence of DIF on 5 of 11 tests in ARIC-NCS but only 2 of 14 tests in BLSA. For the analysis of vision impairment, there was a smaller difference in the number of DIF findings between the samples, as we found evidence of DIF on 2 of 11 items in ARIC and 2 of 14 items in BLSA. However, the measurement of vision in ARIC-NCS entailed selection into the EyeDOC sub-study, which had additional criteria, including requiring participants to score at least a 23/30 in the Jackson, MS site or at least a 25/30 in the Washington County, MD site on the Mini-Mental State Examination (MMSE).

Items that rely strongly on vision, including the Trail-Making Test, Parts A & B, and clock drawing, did not show evidence of DIF (although digit symbol substitution, which relies on vision did have poorer discrimination). Similarly, items that rely strongly on hearing, i.e., the digit span backwards test and the logical memory test, did not show evidence of DIF. This indicates that, contrary to our hypothesis, among people who had levels of vision and hearing function sufficient to validly complete cognitive testing, there may not be difficulty with test administration due to vision and hearing impairments, respectively. It is possible that participants in BLSA and ARIC-NCS are a selected sample due to having volunteer (BLSA), selection based on cognition (EyeDOC sub-study) and survival bias to measurement of sensory characteristics (BLSA and ARIC-NCS), and that these individuals are better able to compensate for their impairment as compared to the general population of those with hearing and vision impairments.

We observed DIF on vision in tests whose administration do not rely on vision, and DIF on hearing in tests whose administration do not rely on hearing. Though this finding is difficult to explain, if verified, it might be consistent with the hypothesis that cognitive load may be playing a role in mediating the effect of sensory impairment on cognitive functioning. Previous studies focusing on hearing impairment have shown that in settings where auditory perception is made more difficult, additional cognitive resources are recruited to help decode sound (Rönnberg et al., 2013; Tun, McCoy, & Wingfield, 2009). The Framework for Understanding Effortful Listening (FUEL) draws on this prior evidence and posits that both motivation and demand, which may be affected by sensory impairments, combine to influence perceived effort (Pichora-Fuller et al., 2016). Although there is less evidence from studies directly assessing the impact of degraded visual stimuli on cognitive processes and cognitive load, a similar mechanism could be hypothesized, whereby additional resources are recruited to understand visual cues and fewer cognitive resources are available (Roberts & Allen, 2016).

The effect of increased cognitive load due to either vision loss or effortful listening may depend on individual-level characteristics, such as working memory capacity (Nitsan, Wingfield, Lavie, & Ben-David, 2019). Our observation of a larger number of DIF findings in less healthy and more cognitively impaired samples may indicate that the effect of

cognitive load may play a larger role in these populations. Future research, investigating differences in the effect of cognitive load by characteristics of how individuals perform on encoding trials for memory tasks, or by individual comorbidities and health factors, may provide further insights into the mechanisms that underlie the effects of cognitive load.

Despite finding evidence of DIF on a number of cognitive items, there was little evidence of salient DIF. This indicates the measurement of cognition among most individuals who were able to complete cognitive testing might not be biased due to sensory impairments in these samples. This finding strengthens previous evidence linking hearing and vision impairment with cognitive impairment by suggesting these relationships are likely not explained by differential measurement error in the measurement of cognition in individuals with hearing and vision impairments.

Interestingly, we found that the Boston Naming Test, a visually-presented measure of object naming ability, appeared to be easier for those with hearing impairment in ARIC-NCS and easier for those with vision impairment in both BLSA and ARIC-NCS. It is possible that individuals with sensory impairments have compensated for these impairments through different mechanisms. Individuals with hearing impairments may rely more on vision, thus helping them on a visual task like the Boston Naming Test. However, this finding was not consistent across any other visual tasks. It is unexpected that individuals with vision impairment would perform better on this visual task. Even though we approximately replicated this finding in two independent samples, future research should seek to explain and further replicate this finding beyond BLSA and ARIC-NCS.

This study has limitations. First, our samples were restricted to individuals who completed cognitive testing in BLSA and ARIC-NCS. Although we showed that sensory impairment likely does not meaningfully bias measures of cognition among those who complete testing, if missingness in cognitive testing is related to sensory impairment and cognitive ability, the estimation of cognition without accounting for systematic missingness could still be subject to selection bias. Second, there are a number of different methods for evaluating DIF, and different methods test slightly different hypotheses and can lead to different conclusions (Liu, Yin, Xin, Shao, & Yuan, 2019; Penfield & Lam, 2000). We chose to use alignment analysis to test for DIF in our sample, as alignment analysis easily allows for the testing of DIF by various parameters. This statistical approach is appropriate to the questions that we sought to address (Asparouhov & Muthén, 2014). Fourth, the confirmatory factor analysis models used in this study make a missing at random (MAR) assumption (Muthén, Kaplan, & Hollis, 1987). This assumption requires that missingness in the sample must be able to be explained by the other variables included in the analysis. However, this assumption may be violated if individuals with sensory impairment have missing values for the tests that are most affected by their impairments, as in this case their non-missing data may not fully reflect their underlying cognitive abilities. Ongoing work will further investigate the potential ramifications of data missing not at random in the detection of differential item functioning. Fifth, as the framework for testing for DIF involves a large number of hypothesis tests (for each discrimination and threshold parameter), there is a risk of multiple comparisons. However, we did not apply a multiple comparisons correction, as the goal of the paper was to examine the patterns of findings across the range of tests examined, rather

than to interpret individual findings. To avoid interpreting findings that may have been due to chance, we focused on instances where there were multiple significant adjacent thresholds or significant thresholds at the ends of the distribution of thresholds. We found that overall, of the 286 tests of DIF conducted there were 29 significant findings (10.1%), greater than double the 5% level that would be expected due to random chance. Sixth, if most items in a battery have DIF, the alignment algorithm may interpret these items as representative of the true relationship between sensory impairment and cognition; in such a setting, items that demonstrate the true underlying relationship might be erroneously identified as having DIF. We suspect this may potentially explain some of our unexpected findings in which DIF favors the impaired group. Future research is needed to confirm and further explore and disentangle potential effects as well as resulting implications.

In this study, we report evidence of DIF by hearing impairment for items that do not rely on hearing and evidence of DIF by vision impairment for items that do not rely on vision, which could be consistent with the cognitive load hypothesis. Despite finding DIF in these samples, there was no evidence of salient DIF. Therefore, in well-conducted studies where neurocognitive testing is done in a quiet, face-to-face setting, sensory impairments may not impact the measurement of latent cognitive ability, among individuals who complete cognitive testing. Future research should further explore the potential effects of selection bias related to missing cognitive testing data on investigations of the association between sensory impairments and cognition. Additionally, studies in other samples are needed, as we found differences in DIF by sensory impairments between the BLSA and ARIC-NCS samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Funding:

References

Alzheimer's Association. (2019). 2019 Alzheimer's Disease Facts and Figures (No. 15(3); pp. 321–387). Retrieved from https://www.alz.org/media/documents/alzheimers-facts-and-figures-2019-r.pdf

Asparouhov T, & Muthén B (2014). Multiple-Group Factor Analysis Alignment. Structural Equation Modeling: A Multidisciplinary Journal, 21(4), 495–508. 10.1080/10705511.2014.919210

Benton AL (1962). The visual retention test as a constructional praxis task. Confinia Neurologica, 22, 141–155. 10.1159/000104348 [PubMed: 13967555]

Benton AL, & Hamsher KD (1976). Multilingual examination. Iowa City, IA: University of Iowa.

Blackburn HL, & Benton AL (1957). Revised administration and scoring of the digit span test. Journal of Consulting Psychology, 21(2), 139. [PubMed: 13416432]

Bollen KA (1989). Confirmatory Factor Analysis. In Structural Equations with Latent Variables (pp. 226–318). John Wiley & Sons, Ltd. 10.1002/9781118619179.ch7

Brookmeyer R, Johnson E, Ziegler-Graham K, & Arrighi HM (2007). Forecasting the global burden of Alzheimer's disease. Alzheimer's & Dementia, 3(3), 186–191. 10.1016/j.jalz.2007.04.381

Cella D, Eton DT, Lai J-S, Peterman AH, & Merkel DE (2002). Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. Journal of Pain and Symptom Management, 24(6), 547–561. 10.1016/s0885-3924(02)00529-8 [PubMed: 12551804]

Chen SP, Bhattacharya J, & Pershing S (2017). Association of Vision Loss With Cognition in Older Adults. JAMA Ophthalmology, 135(9), 963–970. 10.1001/jamaophthalmol.2017.2838 [PubMed: 28817745]

Deal JA, Betz J, Yaffe K, Harris T, Purchase-Helzner E, Satterfield S, … Lin FR (2017). Hearing Impairment and Incident Dementia and Cognitive Decline in Older Adults: The Health ABC Study. The Journals of Gerontology: Series A, 72(5), 703–709. 10.1093/gerona/glw069

Delis D, Kramer J, Kaplan E, & Thompkins B (1987). CVLT: California Verbal Learning Test– Adult Version: Manual. San Antonio, TX: Psychological Corporation.

Dupuis K, Pichora-Fuller MK, Chasteen AL, Marchuk V, Singh G, & Smith SL (2015). Effects of hearing and vision impairments on the Montreal Cognitive Assessment. Aging, Neuropsychology, and Cognition, 22(4), 413–437. 10.1080/13825585.2014.968084

Ehrlich JR, & Langa KM (2019). Potential for dementia prevention in Latin America and Africa based on population-attributable fraction estimates. The Lancet Global Health, 7(10), e1323. 10.1016/S2214-109X(19)30329-8 [PubMed: 31537360]

Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, … Wood WC (2004). A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. Journal of Clinical Epidemiology, 57(9), 898–910. 10.1016/j.jclinepi.2004.01.012 [PubMed: 15504633]

Ferris III FL, Kassoff A, Bresnick GH, & Bailey I (1982). New visual acuity charts for clinical research. American Journal of Ophthalmology, 94(1), 91–96. [PubMed: 7091289]

Goman AM, & Lin FR (2016). Prevalence of Hearing Loss by Severity in the United States. American Journal of Public Health, 106(10), 1820–1822. 10.2105/AJPH.2016.303299 [PubMed: 27552261]

Gussekloo J, de Craen AJM, Oduber C, van Boxtel MPJ, & Westendorp RGJ (2005). Sensory Impairment and Cognitive Functioning in Oldest-Old Subjects: The Leiden 85+ Study. The American Journal of Geriatric Psychiatry, 13(9), 781–786. 10.1097/00019442-200509000-00006 [PubMed: 16166407]

Hebert LE, Weuve J, Scherr PA, & Evans DA (2013). Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. Neurology, 80(19), 1778–1783. 10.1212/WNL.0b013e31828726f5 [PubMed: 23390181]

Jones RN (2006). Identification of Measurement Differences Between English and Spanish Language Versions of the Mini-Mental State Examination: Detecting Differential Item Functioning Using MIMIC Modeling. Medical Care, 44(11), S124–S133. JSTOR. Retrieved from JSTOR. [PubMed: 17060819]

Jones RN, & Gallo JJ (2002). Education and Sex Differences in the Mini-Mental State ExaminationEffects of Differential Item Functioning. The Journals of Gerontology: Series B, 57(6), P548–P558. 10.1093/geronb/57.6.P548

Lin FR, Ferrucci L, An Y, Goh JO, Doshi J, Metter EJ, … Resnick SM (2014). Association of hearing impairment with brain volume changes in older adults. NeuroImage, 90, 84–92. 10.1016/j.neuroimage.2013.12.059 [PubMed: 24412398]

Lin Frank R., & Albert M (2014). Hearing loss and dementia – who is listening? Aging & Mental Health, 18(6), 671–673. 10.1080/13607863.2014.915924 [PubMed: 24875093]

Lin Frank R., Ferrucci L, Metter EJ, An Y, Zonderman AB, & Resnick SM (2011). Hearing Loss and Cognition in the Baltimore Longitudinal Study of Aging. Neuropsychology, 25(6), 763–770. 10.1037/a0024238 [PubMed: 21728425]

Lin Frank R., Niparko JK, & Ferrucci L (2011). Hearing Loss Prevalence in the United States. Archives of Internal Medicine, 171(20), 1851–1853. 10.1001/archinternmed.2011.506 [PubMed: 22083573]

Lin H, Zhang L, Lin D, Chen W, Zhu Y, Chen C, … Chen W (2018). Visual Restoration after Cataract Surgery Promotes Functional and Structural Brain Recovery. EBioMedicine, 30, 52–61. 10.1016/j.ebiom.2018.03.002 [PubMed: 29548900]

Lin MY, Gutierrez PR, Stone KL, Yaffe K, Ensrud KE, Fink HA, … Mangione CM (2004). Vision Impairment and Combined Vision and Hearing Impairment Predict Cognitive and Functional Decline in Older Women. Journal of the American Geriatrics Society, 52(12), 1996–2002. 10.1111/j.1532-5415.2004.52554.x [PubMed: 15571533]

Liu Y, Yin H, Xin T, Shao L, & Yuan L (2019). A Comparison of Differential Item Functioning Detection Methods in Cognitive Diagnostic Models. Frontiers in Psychology, 10. 10.3389/fpsyg.2019.01137

Lord F (1980). Applications of Item Response Theory to Practical Testing Problems. New York: Routledge: Taylor & Fracis Group.

Morales LS, Flowers C, Gutierrez P, Kleinman M, & Teresi JA (2006). Item and Scale Differential Functioning of the Mini-Mental State Exam Assessed Using the Differential Item and Test Functioning (DFIT) Framework. Medical Care, 44(11 Suppl 3), S143–S151. 10.1097/01.mlr.0000245141.70946.29 [PubMed: 17060821]

Muthén B, & Asparouhov T (2014). IRT studies of many groups: The alignment method. Frontiers in Psychology, 5. 10.3389/fpsyg.2014.00978

Muthén B, Kaplan D, & Hollis M (1987). On structural equation modeling with data that are not missing completely at random. Psychometrika, 52(3), 431–462. 10.1007/BF02294365

Nitsan G, Wingfield A, Lavie L, & Ben-David BM (2019). Differences in Working Memory Capacity Affect Online Spoken Word Recognition: Evidence From Eye Movements. Trends in Hearing, 23, 2331216519839624. 10.1177/2331216519839624 [PubMed: 31010398]

Peelle JE, Troiani V, Grossman M, & Wingfield A (2011). Hearing Loss in Older Adults Affects Neural Systems Supporting Speech Comprehension. Journal of Neuroscience, 31(35), 12638–12643. 10.1523/JNEUROSCI.2559-11.2011 [PubMed: 21880924]

Penfield RD, & Lam TCM (2000). Assessing Differential Item Functioning in Performance Assessment: Review and Recommendations. Educational Measurement: Issues and Practice, 19(3), 5–15. 10.1111/j.1745-3992.2000.tb00033.x

Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, Humes LE, … Wingfield A (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). Ear and Hearing, 37, 5S. 10.1097/AUD.0000000000000312 [PubMed: 27355771]

Pickles J (2012). An introduction to the physiology of hearing (4th ed.). Bradford: Emerald Group Publishing Limited.

Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, … Rodgers WL (2007). Prevalence of dementia in the United States: The aging, demographics, and memory study. Neuroepidemiology, 29(1–2), 125–132. [PubMed: 17975326]

Prevention of Blindness and Deafness Grades of Hearing Impairment. (n.d.). Retrieved February 25, 2020, from WHO Website website: http://www.who.int/deafness/hearing_impairment_grades/en/

Reitan RM (1986). Trail Making Test: Manual for administration and scoring: Reitan Neuropsychological Laboratory: Tucson. AZ.

Reitan Ralph M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. Perceptual and Motor Skills, 8(3), 271–276. 10.2466/pms.1958.8.3.271

Roberts KL, & Allen HA (2016). Perception and Cognition in the Ageing Brain: A Brief Review of the Short- and Long-Term Links between Perceptual and Cognitive Decline. Frontiers in Aging Neuroscience, 8. 10.3389/fnagi.2016.00039

Rogers MAM, & Langa KM (2010). Untreated Poor Vision: A Contributing Factor to Late-Life Dementia. American Journal of Epidemiology, 171(6), 728–735. 10.1093/aje/kwp453 [PubMed: 20150357]

Rönnberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, … Rudner M (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. Frontiers in Systems Neuroscience, 7. 10.3389/fnsys.2013.00031

Rucker DD, McShane BB, & Preacher KJ (2015). A researcher's guide to regression, discretization, and median splits of continuous variables. Journal of Consumer Psychology, 25(4), 666–678. 10.1016/j.jcps.2015.04.004

Ryan JJ, & Lopez SJ (2001). Wechsler Adult Intelligence Scale-III. In Dorfman WI & Hersen M (Eds.), Understanding Psychological Assessment (pp. 19–42). Boston, MA: Springer US. 10.1007/978-1-4615-1185-4_2

Shock N, Greulich R, Arenberg D, Costa P, Lakatta E, & Tobin J (1984). Normal Human Aging: The Baltimore Longitudinal Study of Aging. Washington D.C.: US Department of Health and Human Services.

Smith A (1968). The symbol-digit modalities test: A neuropsychologic test of learning and other cerebral disorders. In: Seattle, WA: Special Child Publications; 1968: 83–91. In Learning Disorders (pp. 83–91). Seattle, WA: Special Child Publications.

Swenor BK, Ramulu PY, Willis JR, Friedman D, & Lin FR (2013). The Prevalence of Concurrent Hearing and Vision Impairment in the United States. JAMA Internal Medicine, 173(4), 312–313. 10.1001/jamainternmed.2013.1880 [PubMed: 23338042]

Taljaard DS, Olaithe M, Brennan-Jones CG, Eikelboom RH, & Bucks RS (2016). The relationship between hearing impairment and cognitive function: A meta-analysis in adults. Clinical Otolaryngology, 41(6), 718–729. 10.1111/coa.12607 [PubMed: 26670203]

Tay T, Wang JJ, Kifley A, Lindley R, Newall P, & Mitchell P (2006). Sensory and Cognitive Association in Older Persons: Findings from an Older Australian Population. Gerontology, 52(6), 386–394. 10.1159/000095129 [PubMed: 16921251]

The ARIC investigators. (1989). The Atherosclerosis Risk in Communities (ARIC) Study: Design and objectives. American Journal of Epidemiology, 129(4), 687–702. [PubMed: 2646917]

Tun PA, McCoy S, & Wingfield A (2009). Aging, hearing acuity, and the attentional costs of effortful listening. Psychology and Aging, 24(3), 761–766. 10.1037/a0014802 [PubMed: 19739934]

Vitale S, Cotch MF, & Sperduto RD (2006). Prevalence of Visual Impairment in the United States. JAMA, 295(18), 2158–2163. 10.1001/jama.295.18.2158 [PubMed: 16684986]

Wechsler D (1987). The Wechsler Memory Scale—Revised. San Antonio, Texas: Psychological Corp, Harcourt.

Williams BW, Mack W, & Henderson VW (1989). Boston naming test in Alzheimer's disease. Neuropsychologia, 27(8), 1073–1079. 10.1016/0028-3932(89)90186-3 [PubMed: 2797414]

Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, & Cella D (2005). Using Multiple Anchor- and Distribution-Based Estimates to Evaluate Clinically Meaningful Change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) Instrument. Value in Health, 8(2), 117–127. 10.1111/j.1524-4733.2005.08202.x [PubMed: 15804320]

Zheng DD, Swenor BK, Christ SL, West SK, Lam BL, & Lee DJ (2018). Longitudinal Associations Between Visual Impairment and Cognitive Functioning: The Salisbury Eye Evaluation Study. JAMA Ophthalmology, 136(9), 989–995. 10.1001/jamaophthalmol.2018.2493 [PubMed: 29555805]
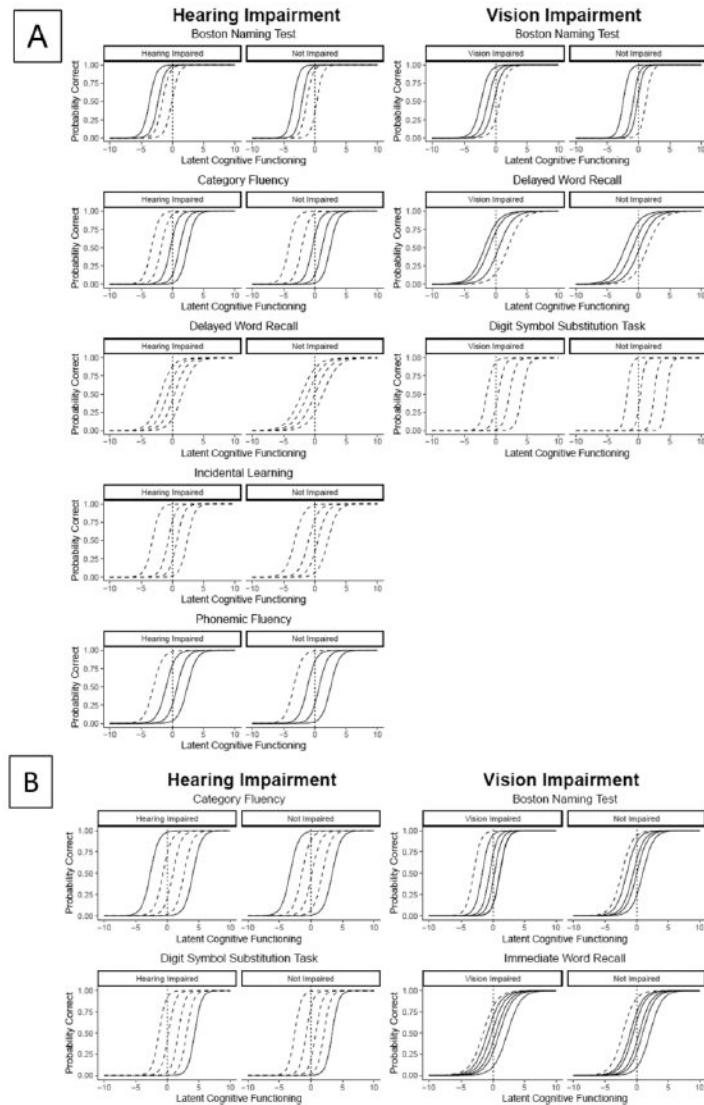
Figure 1.

Item characteristic curves (ICCs) for items with significant differential item functioning (DIF) in the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS) [A] and the Baltimore Longitudinal Study of Aging (BLSA) [B]. ICCs describe for each threshold of each item the probability of achieving that threshold of that item given a particular level of latent cognitive functioning. For example, DIF on thresholds is evident in the delayed word recall task for those with and without vision impairment in BLSA. The leftmost curve is shifted further right in the impaired group as compared to the group without vision impairment, indicating that the item is easier in the unimpaired group and harder in the group with vision impairment. DIF on discrimination can be visualized by inspecting the ICCs for the incidental learning item for those with and without hearing impairment in ARIC-NCS. The curves in the unimpaired group are more slanted, and have a lower slope, indicating that the relationship between the item and the latent cognitive trait is weaker in the unimpaired group as compared to the group with hearing impairment. Dotted lines indicate statistically significant DIF at the $p < 0.05$ level.
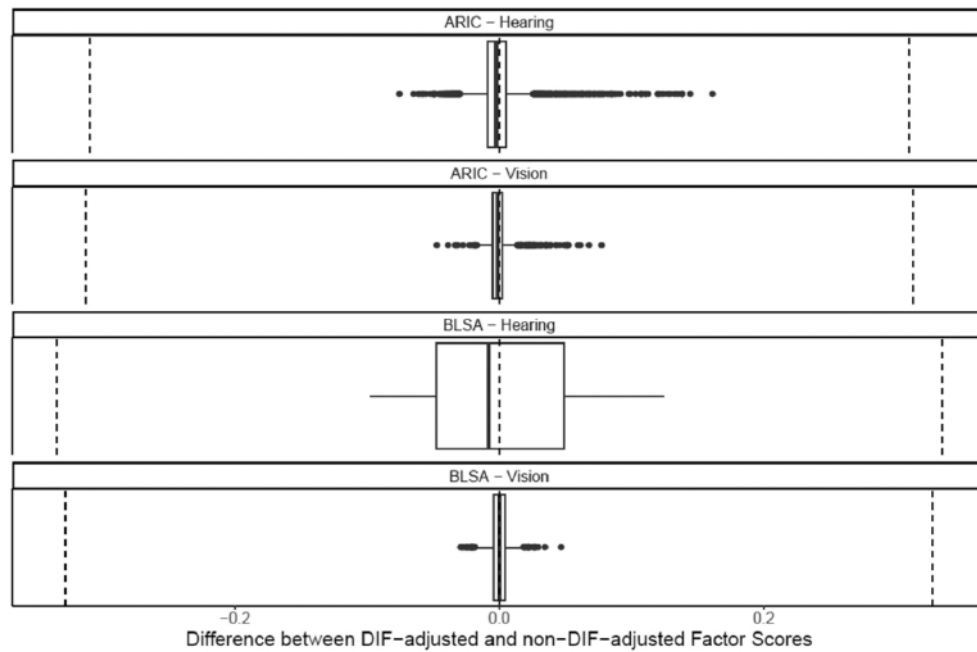
Figure 2.
Differences in differential item functioning (DIF)-adjusted and non-DIF-adjusted factor scores to evaluate salient DIF by Hearing and Vision Impairment in the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS) and the Baltimore Longitudinal Study of Aging (BLSA). The box and whisker plots show the distribution of individual-level differences in factor scores (a measure of latent cognitive ability) from a model adjusted for DIF and a model ignoring DIF. Dotted vertical lines indicate 0 as well as ± 1 SE of measurement, which is a threshold indicating the presences of clinically meaningful DIF.

Table 1.

Cognitive tests included in the Atherosclerosis Risk in Communities Study Neurocognitive Study (ARIC-NCS) and Baltimore Longitudinal Study of Aging (BLSA) and their primary sensory modalities

Cognitive Test	ARIC	BLSA	Primary Sensory Modality
Language			
Category fluency	X	X	Hearing
Phonemic fluency	X	X	Hearing
Boston Naming Task	X	X	Vision
Memory			
Figural memory		X	Vision
Delayed free recall	X	X	Both
Logical memory	X		Hearing
Incidental learning	X		Vision
Attention			
Trail making test A	X	X	Vision
WAIS-R Digits Forward		X	Hearing
Executive			
WAIS-R Digits Backwards	X	X	Hearing
Digit Symbol Substitution	X	X	Vision
Trail making test B	X	X	Vision
Similarities Task		X	Hearing
Card rotation Task		X	Vision
Visuospatial Ability			
Clock Drawing		X	Vision

Category fluency was animal naming in ARIC, but the mean of three trials (fruits, animals, vegetables) in BLSA, Phonemic fluency was the sum of three trials in ARIC, but the mean of three trials in BLSA, Boston Naming Task 30 item version was administered in ARIC, Boston Naming Task 60 item version was administered in BLSA. Delayed free recall was from the California Verbal Learning Test in BLSA, but the ARIC designed and used its own free recall test.

Table 2.

Demographic characteristics and cognitive test scores in the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS) and the Baltimore Longitudinal Study of Aging (BLSA)

	ARIC-NCS	BLSA
Participants in Hearing Sample	N = 3591	N = 671
Participants in the Vision Sample	N = 1051	N = 703
Hearing Impairment - Impaired - N (%)	2622 (73.0)	354 (52.2)
Vision Impairment - Impaired - N (%)	191 (18.2)	180 (25.2)
Age - Mean (Standard Deviation - SD)	79.8 (4.7)	74.7 (9.0)
Education: Less than HS - N (%)	437 (12.0)	2 (0.3)
Education: HS or equivalent - N (%)	1499 (41.3)	34 (4.6)
Education: Beyond HS - N (%)	1692 (46.6)	696 (95.1)
Race - Black - N (%)	869 (24.0)	249 (34.0)
Sex - Female - N (%)	2172 (59.8)	407 (55.6)
Cognitive Items (Raw Scores) - Mean (SD)		
Benton Visual Retention Task[†]		10.8 (5.5)
Boston Naming Test (30 Item)[†]	25.4 (4.7)	
Boston Naming Test (60 Item)[†]		53.6 (6.5)
Card Rotations[†]		82.7 (40.3)
Category Fluency (Animals)[‡]	15.5 (5.0)	
Category Fluency (Mean of 3 Categories)[‡]		15.3 (4.8)
Clock 11:10[†]		8.9 (1.4)
Clock 3:35[†]		8.5 (1.3)
Delayed Word Recall	5.2 (1.8)	
Delayed Word Recall (CVLT)[‡]		10.2 (3.7)
Digit Span Backwards[‡]	5.5 (1.9)	6.9 (2.2)
Digit Span Forwards[‡]		8.0 (2.2)
Digit Symbol Substitution Task[†]	36.9 (11.7)	40.3 (11.9)
Incidental Learning[†]	6.0 (1.7)	
Logical Memory Story 1[‡]	21.9 (7.4)	
Logical Memory Story 2[‡]	17.7 (8.1)	
Phonemic Fluency (Mean of 3 Trials)[‡]		14.5 (5.2)
Phonemic Fluency (Sum of 3 Trials)[‡]	32.4 (12.3)	
Similarities[‡]		21.1 (3.6)
Trail-Making Test A[†]	51.6 (28.0)	35.9 (24.2)
Trail-Making Test B[†]	132.1 (59.4)	88.6 (46.5)

†Items whose assessments primarily relied solely on vision are indicated with.

‡Items whose assessments relied primarily solely on hearing are indicated with.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Parameter Differences by Vision Impairment from Alignment Models in the Baltimore Longitudinal Study of Aging (BLSA) and Atherosclerosis Risk in Communities Study the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS)

Sample	Cognitive Test	Discrimination	Thresholds					
			1	2	3	4	5	6
BLSA (N=703)	Benton Visual Retention Test	0.25 (0.31)	0.06 (0.28)	0.20 (0.23)	0.30 (0.22)	0.31 (0.24)	0.21 (0.31)	0.65 (0.40)
	Boston Naming Test	0.53 (0.33)	−0.74 (0.35)*	−0.15 (0.27)	0.02 (0.23)	0.21 (0.22)	0.10 (0.24)	
	Card Rotations	−0.20 (0.39)	0.03 (0.28)	0.02 (0.20)	−0.01 (0.22)	0.20 (0.36)	−0.53 (0.42)	
	Category Fluency	0.05 (0.23)	0.35 (0.33)	0.09 (0.24)	0.23 (0.27)	−0.16 (0.34)	−0.17 (0.49)	
	Clock 11:10	0.16 (0.31)	−0.12 (0.27)	−0.20 (0.22)	−0.06 (0.22)			
	Clock 3:35	0.17 (0.29)	0.22 (0.36)	0.11 (0.24)	−0.09 (0.21)	−0.03 (0.30)		
	Delayed Word Recall	−0.12 (0.23)	0.62 (0.28)*	0.15 (0.22)	0.32 (0.21)	0.04 (0.21)	0.06 (0.23)	0.17 (0.30)
	Digit Span Backwards	0.16 (0.24)	−0.48 (0.26)	−0.20 (0.22)	−0.17 (0.23)	−0.19 (0.37)	−0.58 (0.47)	
	Digit Span Forwards	−0.05 (0.17)	−0.18 (0.28)	−0.27 (0.22)	−0.19 (0.21)	−0.23 (0.22)	−0.52 (0.24)*	−0.44 (0.28)
	Digit Symbol Substitution Task	−0.25 (0.43)	0.17 (0.27)	0.10 (0.23)	0.07 (0.27)	0.01 (0.35)	−0.32 (0.51)	
	Phonemic Fluency	0.04 (0.15)	−0.34 (0.27)	−0.41 (0.21)	−0.40 (0.25)	0.01 (0.40)		
	Similarities	0.34 (0.27)	−0.52 (0.33)	0.05 (0.23)	−0.07 (0.22)	0.08 (0.21)	0.22 (0.29)	
	Trail-Making Test A	−0.45 (0.46)	0.37 (0.33)	0.08 (0.22)	−0.14 (0.27)	−0.16 (0.39)	0.29 (0.55)	
	Trail-Making Test B	−0.19 (0.54)	0.01 (0.38)	−0.05 (0.26)	0.04 (0.27)	0.08 (0.39)	−0.09 (0.52)	−0.31 (0.64)
ARIC (N=1051)	Boston Naming Test	−0.62 (0.35)	0.05 (0.29)	−0.41 (0.24)	−0.17 (0.23)	−0.55 (0.24)*		
	Category Fluency	−0.05 (0.20)	−0.26 (0.40)	−0.22 (0.27)	−0.20 (0.22)	0.03 (0.25)	−0.54 (0.31)	
	Delayed Word Recall	0.03 (0.15)	0.34 (0.24)	−0.14 (0.20)	0.27 (0.19)	0.50 (0.24)*		
	Digit Span Backwards	0.19 (0.27)	−0.13 (0.26)	−0.11 (0.22)	0.10 (0.22)	−0.22 (0.25)	0.39 (0.54)	
	Digit Symbol Substitution Task	−0.93 (0.42)*	0.38 (0.28)	0.08 (0.25)	−0.34 (0.34)	−0.31 (0.52)		
	Incidental Learning	0.10 (0.27)	−0.06 (0.34)	0.01 (0.21)	0.17 (0.24)	0.25 (0.35)		
	Logical Memory 1	0.93 (0.59)	−0.36 (0.40)	0.23 (0.27)	0.53 (0.44)	0.70 (0.73)		
	Logical Memory 2	1.05 (0.61)	−0.44 (0.43)	0.04 (0.26)	0.40 (0.38)	0.50 (0.59)	0.80 (0.88)	
	Phonemic Fluency	0.01 (0.17)	0.10 (0.30)	0.07 (0.20)	0.31 (0.23)	0.16 (0.31)		
	Trail-Making Test A	−0.64 (0.39)	−0.08 (0.30)	−0.06 (0.21)	−0.26 (0.31)	0.21 (0.53)		
	Trail-Making Test B	−0.29 (0.53)	−0.16 (0.28)	0.36 (0.34)	0.09 (0.42)	−0.01 (0.62)	−0.15 (0.76)	

Data show parameter differences (SE), parameter differences are the parameter value in the impaired group – the parameter value in the unimpaired group, standardized effect sizes can be calculated by dividing the parameter difference by the standard error. We found significant differences in parameters indicating that delayed word recall in BLSA and ARIC-NCS were harder for those with vision impairment. The Boston naming test was easier in both BLSA and ARIC- NCS for those with vision impairment. The digit symbol substitution task was less strongly related to cognition among those with vision impairment in the ARIC-NCS sample.

*
= p < 0.05

**
= p < 0.01

Table 4.

Parameter Differences by Hearing Impairment from Alignment Models in the Baltimore Longitudinal Study of Aging (BLSA) and the Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS)

Sample	Cognitive Test	Discrimination	Thresholds					
			1	2	3	4	5	6
BLSA (N=671)	Benton Visual Retention Test	0.11 (0.24)	0.27 (0.35)	0.32 (0.30)	0.20 (0.28)	0.25 (0.28)	0.12 (0.31)	0.41 (0.35)
	Boston Naming Test	0.15 (0.27)	-0.47 (0.41)	-0.55 (0.34)	-0.33 (0.30)	-0.31 (0.29)	-0.44 (0.28)	
	Card Rotations	-0.04 (0.20)	-0.45 (0.38)	-0.20 (0.30)	0.04 (0.28)	-0.10 (0.31)	-0.59 (0.40)	
	Category Fluency	0.04 (0.25)	0.63 (0.41)	0.75 (0.32)*	0.72 (0.32)*	0.87 (0.39)*	0.71 (0.50)	
	Clock 11:10	-0.18 (0.27)	-0.34 (0.30)	-0.31 (0.27)	-0.25 (0.26)			
	Clock 3:35	-0.28 (0.25)	-0.05 (0.34)	0.13 (0.26)	-0.21 (0.25)			
	Delayed Word Recall	0.18 (0.22)	0.55 (0.32)	0.26 (0.27)	0.36 (0.26)	0.38 (0.25)	0.21 (0.26)	0.50 (0.29)
	Digit Span Backwards	0.06 (0.23)	-0.08 (0.23)	-0.19 (0.19)	0.01 (0.17)	-0.12 (0.27)	-0.32 (0.39)	
	Digit Span Forwards	-0.19 (0.20)	0.32 (0.29)	-0.05 (0.18)	-0.15 (0.15)	-0.05 (0.13)	-0.02 (0.13)	0.03 (0.16)
	Digit Symbol Substitution Task	-0.06 (0.24)	1.25 (0.37)**	0.91 (0.37)*	0.98 (0.41)*	1.01 (0.46)*	1.00 (0.56)	
	Phonemic Fluency	-0.25 (0.23)	-0.44 (0.30)	-0.41 (0.25)	-0.42 (0.28)	-0.15 (0.36)		
	Similarities	-0.14 (0.24)	-0.38 (0.34)	-0.33 (0.28)	-0.40 (0.27)	-0.48 (0.26)	-0.47 (0.29)	
	Trail-Making Test A	0.00 (0.21)	0.61 (0.41)	0.26 (0.33)	0.17 (0.37)	0.22 (0.43)	0.81 (0.50)	
	Trail-Making Test B	0.10 (0.33)	0.24 (0.47)	-0.06 (0.41)	0.35 (0.43)	0.71 (0.51)	0.73 (0.58)	1.13 (0.67)
ARIC (N=3591)	Boston Naming Test	-0.29 (0.15)	-0.21 (0.20)	-0.21 (0.14)	-0.28 (0.12)*	-0.36 (0.12)**		
	Category Fluency	0.00 (0.09)	0.85 (0.23)**	0.34 (0.14)*	0.12 (0.10)	0.05 (0.10)	0.01 (0.13)	
	Delayed Word Recall	0.20 (0.08)*	0.26 (0.12)*	0.40 (0.09)**	0.38 (0.09)**	0.33 (0.10)**		
	Digit Span Backwards	-0.08 (0.09)	-0.21 (0.13)	-0.11 (0.09)	-0.01 (0.08)	-0.05 (0.10)	0.07 (0.16)	
	Digit Symbol Substitution Task	0.17 (0.19)	-0.14 (0.19)	0.20 (0.13)	-0.07 (0.13)	0.13 (0.19)		
	Incidental Learning	0.35 (0.11)**	0.04 (0.17)	0.35 (0.10)**	0.31 (0.10)**	0.33 (0.14)*		
	Logical Memory 1	0.03 (0.11)	0.23 (0.16)	0.11 (0.10)	-0.16 (0.10)	-0.04 (0.16)		
	Logical Memory 2	-0.02 (0.11)	0.28 (0.16)	0.29 (0.11)*	0.03 (0.09)	0.00 (0.12)	-0.17 (0.22)	
	Phonemic Fluency	-0.10 (0.10)	0.37 (0.17)*	0.19 (0.11)	0.11 (0.10)	-0.07 (0.13)		
	Trail-Making Test A	-0.06 (0.15)	-0.26 (0.15)	-0.28 (0.11)**	-0.06 (0.12)	-0.18 (0.18)		
	Trail-Making Test B	0.06 (0.19)	-0.24 (0.14)	-0.17 (0.12)	-0.12 (0.15)	-0.19 (0.21)	-0.07 (0.28)	

Data show parameter differences (SE), parameter differences are the parameter value in the impaired group – the parameter value in the unimpaired group, standardized effect sizes can be calculated by dividing the parameter difference by the standard error. We found significant differences such that the category fluency test, and the digit symbol substitution task in BLSA, and the delayed word recall test, the incidental learning test, and the phonemic fluency test in ARIC- NCS were harder for those with hearing impairment. The Boston naming test was easier for those with hearing impairment in ARIC-NCS. The delayed word recall and incidental learning tests in ARIC-NCS were more strongly related to cognition in those with hearing impairment.

*
= $p < 0.05$

**
= $p < 0.01$