Received: 27 August 2024 | Revised: 22 January 2025 | Accepted: 8 February 2025

DOI: 10.1002/alz.70097

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Harmonized cognitive performance in an older adult cohort of Vietnamese American immigrants: The VIP study

Brandon E. Gavett[1] | Sarah Tomaszewski Farias[1] | Quyen Q. Tiet[2,3] | Van T. Park[4] | Danielle Harvey[1] | Quyen Vuong[5] | Ladson Hinton[1] | Alka M. Kanaya[4] | Rachel A. Whitmer[1] | Lauren Mai[6] | Oanh L. Meyer[1]

[1]Department of Neurology, University of California Davis School of Medicine, Sacramento, California, USA

[2]Clinical Psychology PhD Program, California School of Professional Psychology at Alliant International University, Emeryville, California, USA

[3]National Center for PTSD, Dissemination and Training Division, VA Palo Alto Health Care System, Menlo Park, California, USA

[4]Department of Medicine, University of California San Francisco, San Francisco, California, USA

[5]International Children Assistance Network (ICAN), Milpitas, California, USA

[6]Zuckerberg San Francisco General Hospital and Trauma Center, San Francisco, California, USA

Correspondence
Brandon Gavett, University of California Davis School of Medicine, 1651 Alhambra Blvd. Suite 200A, Sacramento, CA 95816, USA.
Email: bgavett@ucdavis.edu

Funding information
National Institute on Aging, Grant/Award Numbers: R01AG067541, R24AG063718, P30AG072972

Abstract

INTRODUCTION: Vietnamese Americans represent an understudied population with unique risk factors relevant to cognitive aging. The current study sought to model global cognition in the Vietnamese Insights into Cognitive Aging Program (VIP) study and harmonize ability estimates with the National Alzheimer's Coordinating Center (NACC) Uniform Data Set.

METHODS: Cognitive data from VIP (N = 548) and NACC (N = 15,923) were analyzed using item response theory. Seven common items were assessed for differential item functioning (DIF); items without salient DIF were used to harmonize the cognitive composite score across the two cohorts.

RESULTS: Although five of the seven common items showed evidence of DIF, the magnitude of this DIF was negligible, affecting the factor score estimates of only 12 (2.19%) VIP participants by more than one standard error.

DISCUSSION: Global cognitive functioning can be estimated in Vietnamese American immigrants with minimal bias and psychometrically matched to one of the largest studies of cognitive aging and dementia worldwide.

KEYWORDS
aging, cognition, cultural diversity, item response theory, neuropsychological tests

Highlights

- This is the first known study to model cognition in older Vietnamese Americans.
- Global cognition was harmonized with minimal bias across two diverse cohorts.
- Differential item functioning was found in five of seven items, but the impact was not salient.
- Results create new opportunities to study health disparities in an underrepresented group.

1 | BACKGROUND

Historically, cognitive aging research has included mainly non-Hispanic White English-speaking participants.[1,2] Consequently, our knowledge about Alzheimer's disease and related dementias (ADRD) may not fully represent the breadth of individual differences and life experiences that influence risk and protective factors governing late-life cognitive decline and dementia.[3,4] This knowledge may help improve health and reduce health care disparities among older adults from underrepresented groups.[5–7] For instance, understanding why marked disparities in rates of cognitive impairment exist across different racial and ethnic groups[8] requires recruitment of diverse individuals—many of whom do not speak English—into longitudinal studies of cognitive aging.

One barrier to inclusivity and representativeness has been the lack of validated cognitive assessment instruments for individuals who do not speak English or who lack familiarity with the culture in which these assessments were developed.[9–11] Recruiting diverse groups can lead to important advances toward understanding factors that shape late-life cognitive decline.[12] For instance, Asian American participants are often grouped together when analyzing data by race. But considerable heterogeneity exists among Asian American individuals and across ethnic subgroups, which can influence associations between exposures and outcomes relevant to health disparities.[13,14] Many older Vietnamese Americans have experienced trauma and hardships associated with war and migration that may influence late-life cognitive health in ways that are not currently well understood.[15] War, trauma, and migration have been ubiquitous throughout human history; therefore, understanding how these exposures shape cognitive aging and the impact of and ethnic and cultural heterogeneity in modifying these associations is likely to have continued relevance for current and future generations.[16,17]

Measuring cognitive functioning in people from different language backgrounds is not as simple as translating items into a new language.[18,19] Cognitive abilities—and choices about how to measure them—are influenced by cultural attitudes, beliefs, and practices.[20] This can lead to challenges when comparing cognitive performance across different language and cultural groups. These comparisons can be facilitated by the presence of common assessment outcomes shared between the groups. However, if these common items do not provide equally valid measures of the intended construct in both groups, then fair comparisons may not be possible. When a test item does not have the same measurement properties across groups (e.g., it is easier for one group than another, despite the two groups being matched on the underlying ability being measured), the item is said to have differential item functioning (DIF).[21] DIF has the potential to introduce systematic measurement error that creates misleading results when making group comparisons. Statistical methods, such as item response theory (IRT), can be valuable for identifying items with DIF and promoting fair comparisons across different groups. When using methods such as IRT for scoring cognitive tests, shared DIF-free items can facilitate harmonization of test scores by equating the scale of the metric across the groups being compared.[22]

In the current study, we aim to pursue harmonization of cognitive test scores in a group of elderly Vietnamese Americans. This cohort comes from the Vietnamese Insights into Cognitive Aging Program (VIP) study, which has engaged in targeted recruitment of people who lived in Vietnam and immigrated to the United States during or after the Vietnam war era.[23] The neuropsychological battery in the VIP study was designed to have partial overlap with the neuropsychological battery included in the Uniform Data Set (UDS) version 3, which is administered to research participants at Alzheimer's Disease Research Centers (ADRCs) across the United States. There were four main goals of the current study: (1) to use IRT to model cognitive performance in both groups; (2) to identify overlapping items between the two batteries that could be used as linking items for harmonization purposes; (3) to identify whether any of the candidate linking items contain DIF, which could contraindicate their use as linking items; and (4) to compare the harmonized cognitive factor scores across cohorts.

RESEARCH IN CONTEXT

1. **Systematic review**: A literature review was performed using PubMed and Scopus; this was augmented by an examination of the reference lists from relevant published articles. In recent years, a number of articles have described best practices for pre-statistical and statistical harmonization of cognitive tests across diverse groups. These have been cited and used as exemplars for the current study.
2. **Interpretation**: The results demonstrate that global cognitive ability can be estimated with high reliability in Vietnamese Americans; the ability estimates generated are scaled on the same metric as data from the National Alzheimer's Coordinating Center, a well-characterized longitudinal cohort of older adults.
3. **Future directions**: The results provide a mechanism for making cross-cultural and cross-language comparisons of cognitive performance, which can facilitate future research focusing on how different exposures may influence cognition in this population of older adults who have experienced trauma related to war and migration.

2 | METHODS

The methods described below were influenced by a series of harmonization studies reported by Arce Rentería, Briceño, Gross, Jones, Manly, and colleagues.[24–27]

2.1 | Participants

This study was conducted using two distinct participant cohorts, both of which were established with the overarching goal of studying cognitive aging and the risk of ADRD. The focal group in our analyses was the VIP cohort, composed of Vietnamese Americans who were recruited from two regions in Northern California: Sacramento and Santa Clara. The VIP cohort consists of 548 participants 65 years of age and older, all of whom identified as Vietnamese or Vietnamese American and immigrated to the United States from Vietnam. All spoke Vietnamese as their native language, with varying degrees of English-language proficiency. A detailed description of VIP recruitment methods and sample characteristics has been published recently.[23] Baseline data from all VIP participants were used in the current study. Human subjects research approval in the VIP study was provided by the UC Davis and UC San Francisco Institutional Review Boards; participants provided informed consent and all research protocols were conducted in accordance with the Declaration of Helsinki.

The reference group in our analyses consisted of 15,923 participants from the National Alzheimer's Coordinating Center (NACC) database, which is a collection of participants compiled from 37 ADRCs across the United States. Archival data were obtained with a data use request to NACC. Participant data from the NACC dataset were used in the current study if they were obtained at the first (baseline) visit, if the neuropsychological assessment and Montreal Cognitive Assessment (MoCA) were both administered in English, and if the data were collected using the UDS 3 protocol. In addition, we included only NACC participants who were age 50 or older at the baseline assessment. The NACC UDS 3 protocol has been described in detail.[28,29] NACC data were obtained in de-identified format for the current study; local institutional review boards specific to each ADRC site provided human subjects ethics approval at the time of enrollment.

2.2 | Materials and pre-statistical harmonization

The VIP study is a prospective longitudinal cohort study based in Northern California. Study enrollment began in January 2022 and concluded in November 2023. For the current study, only data from the baseline visit were used.

Prior to data collection, a VIP research team containing cognitive aging researchers, neuropsychologists, and research staff with expertise in cross-cultural assessment—including numerous native Vietnamese speakers—selected and translated a neuropsychological assessment battery for use in this cohort. (See Tiet et al.[30] for a detailed description of the test adaptation process.) This battery was composed largely of tests from the UDS 3 and the Cognitive Abilities Screening Instrument (CASI),[31] plus the World Health Organization—University of California Los Angeles Auditory Verbal Learning Test (AVLT).[32] All examiners, who were fluent in Vietnamese and English, were trained to certification standards by a clinical neuropsychologist and a bilingual clinical psychologist before administering the neuropsychological battery. All tests were conducted in the Vietnamese language and administered using paper and pencil forms, either in the participant's home or at the offices of community-based organizations with whom we partnered as part of participant recruitment and retention. (See Meyer et al.[23] for a detailed description of the VIP study protocol.)

After more than 2 years spent administering this battery to participants and conducting research diagnostic case conferences based on the assessment results, a survey of the research team was conducted to identify items that were expected to perform equivalently across the VIP and NACC cohorts. This survey inquired about items or variables from the VIP battery that were exact duplicates of—or highly similar to—those used in the UDS 3. The research team rated each item on its equivalence between Vietnamese and English, focusing on administration, scoring, interpretation, language, culture, and construct validity (see Table S1).[25] Based on this survey, seven items were chosen for use as potential linking items for harmonization across cohorts: (1) Animal Fluency, (2) Benson Figure Copy, (3) Benson Figure Delayed Recall, (4) Benson Figure Recognition, (5) Number Span Forward, (6) Number Span Backward, and (7) Trail Making Test Part A.

In total, 30 neuropsychological test items were chosen for inclusion from the combined VIP and NACC samples. Of these 30 items, 7 were available in both cohorts and rated as being essentially equivalent by the research team (potential linking items), 12 were unique to VIP, and 11 were unique to NACC. When necessary, items were recoded from their original scales to a polytomous (ordered categorical) scale, with up to 10 categories per item. The recoding scheme required a minimum of 10 responses per category to avoid sparse cells; categories with fewer than 10 responses were collapsed. A description of the items, as well as the recoding scheme used for each item, are provided in Tables S2 and S3.

Because only seven items were available as potential linking items, we were unable to pursue harmonization within specific cognitive domains (e.g., memory, executive functioning) without potentially introducing considerable bias.[33] Instead, we constructed a global cognitive composite factor using a bifactor modeling approach to account for variance attributable to domain-specific cognitive abilities. A path diagram showing the confirmatory factor analysis model used to estimate global cognition can be seen in Figure 1.

2.3 | Statistical harmonization

Statistical analysis was performed in R version 4.3.1.[34] All model fitting was performed using the bfactor function provided by mirt package version 1.41.8.[35] We began by estimating bifactor IRT models separately in the VIP and NACC cohorts to ensure that these models fit the data well in both cohorts. Model fit was judged using the C2 statistic,[36] comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) using standard benchmarks.[37] After establishing good model fit in both cohorts, we saved each item's parameter estimates in an item bank to be used in the harmonization process (fixed parameter calibration).[38] For items unique to the VIP cohort, parameters were saved in the VIP item bank. The parameter estimates from the remaining items (potential linking items and items unique to NACC) were saved in the NACC item bank, consistent with our use of NACC as the reference group.

Before using these results to generate factor scores for the two cohorts, we sought to evaluate whether our assumption that the seven

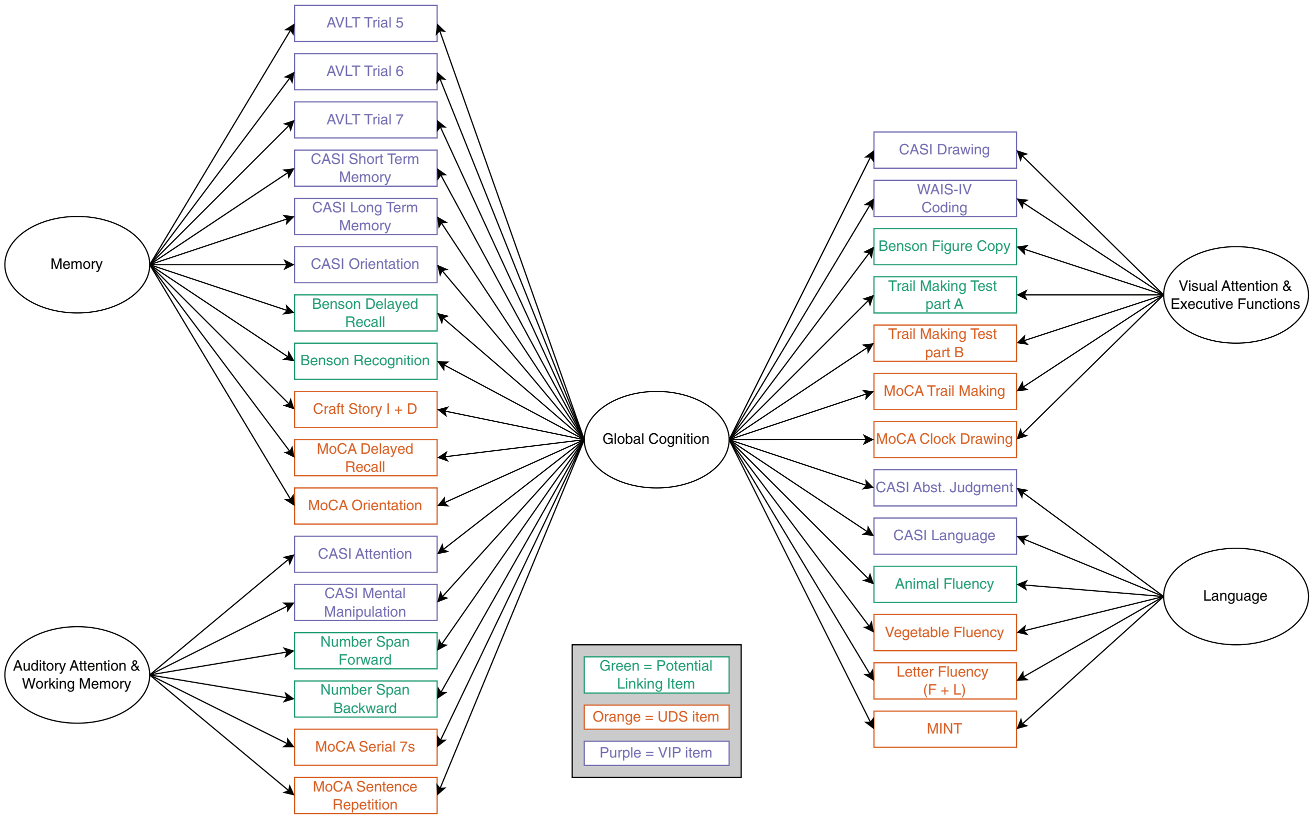

FIGURE 1 Path diagram representing the multidimensional IRT model used to estimate global cognitive ability in the VIP and NACC cohorts. Colors show whether an item was a potential linking item (green), whether it was unique to NACC (orange), or unique to VIP (purple). Abst. Judgment, abstraction and judgment; AVLT, Auditory Verbal Learning Test; CASI, Cognitive Abilities Screening Instrument; Craft Story I + D, Craft Story Immediate plus Delayed recall scores; F + L, letter F plus letter L fluency; IRT, item response theory; MINT, Multilingual Naming Test; MoCA, Montreal Cognitive Assessment; NACC, National Alzheimer's Coordinating Center; VIP, Vietnamese Insights into Cognitive Aging Program; WAIS-IV, Wechsler Adult Intelligence Scale, 4th Edition.

potential linking items were free from DIF might be contradicted by the data.

2.4 | DIF

To perform analysis of DIF for the seven potential linking items, we used multiple groups confirmatory factor analysis. DIF testing requires some items to serve as anchors (items whose parameters are constrained to be equal in both groups) to ensure that results are not contaminated by variance attributable to the latent factor(s). Initially, all seven potential linking items were assumed to be good anchor items because, as described above, our team of experts rated the items as being essentially equivalent in VIP and NACC. We subjected this assumption to empirical verification by systematically varying whether or not the potential linking items were anchor items and determining whether model fit improved when a potential linking item was removed as an anchor item.

The first stage of DIF testing was performed using the all-others-as-anchors approach, which is a backward stepwise selection procedure that builds from a constrained baseline model.[39] Initially a fully constrained model was treated as the reference, where all seven potential linking items were used as anchor items. We then iteratively examined how model fit changed when one of the potential linking items was removed as an anchor while the other potential linking items remained anchors. Change in model fit was compared using the sample-size adjusted Bayesian Information Criterion (SABIC)[40] to identify the item with the strongest evidence of DIF. Other outcomes, including the likelihood ratio chi-square test, Akaike Information Criterion (AIC),[41] Bayesian Information Criterion (BIC),[42] and Hannan-Quinn Criterion (HQC)[43] are provided in the electronic supplement. The model with the largest SABIC value was used to identify the item that demonstrated the most pronounced DIF in each iteration. This item was removed as an anchor item in subsequent iterations. As such, the next iteration generated a new reference model that contained only six anchor items, and so forth. We then performed the same series of model comparisons for each remaining item, removing the item with the strongest evidence of DIF from the set of candidate anchor items. We repeated this process until none of the remaining candidate anchor items showed evidence of DIF.

In the second stage of DIF testing, we re-evaluated the DIF-containing items using an iterative forward stepwise approach. This was done because whether an item demonstrates DIF depends, in part, on the other items used as anchors[44]; as such, items identified as having DIF in the first stage may have been false-positive errors. We began by constructing a reference model whose anchor items were found in the first stage to be DIF-free. We then iteratively added each remaining item to the set of anchor items to determine whether imposing group equality constraints on an item led to a decrement in model fit. As above, we used the SABIC statistic to compare models. If the inclusion of an item as an anchor did not lead to a significant decrement in model fit relative to the reference model, then that item was considered DIF-free and the iterative process continued until no more DIF-free items were found.

Finally, after identifying items with and without DIF, we sought to determine the practical impact of the DIF on estimated factor scores. We generated harmonized factor score estimates for two models: (1) a DIF-adjusted model, where the linking items were limited to those that were DIF-free, and (2) an unadjusted model, where all seven shared items were used as linking items. We compared the DIF-adjusted factor scores to the unadjusted factor scores for each participant. The difference in a participant's two-factor score estimates, divided by the pooled standard error of those estimates, yielded a scaled difference score. The effect of DIF on the participant's unadjusted factor score was considered "salient" if the absolute value of the scaled difference score was greater than or equal to 1 unit of pooled standard error. Subsequently, we calculated the proportion of participants whose unadjusted factor scores were affected by salient DIF.

2.5 | Scoring

After identifying items with and without DIF and assessing the impact of DIF on the unadjusted factor scores, we generated final global factor score estimates, which occurred in two steps.

In the first step, we estimated a multiple groups confirmatory factor analysis model with a combination of fixed and freely estimated parameters. Fixed parameters included those of the linking items (fixed to the banked values from the NACC cohort) and the items unique to a specific cohort (fixed to the banked values from their respective cohort). We also fixed the mean and variance of all latent variables (in both cohorts) to 0 and 1, respectively, with one exception: we freely estimated the mean and variance of the global factor in the VIP cohort.

In the second step, we estimated a model with no free parameters by constraining all item parameters and group hyperparameters (means and variances) to their estimated values from the first step. From this model with no free parameters, we generated harmonized factor scores and standard error estimates for each participant in VIP and NACC, using the expected a posteriori (EAP) method.[45]

2.6 | Group comparisons

We sought to determine how the harmonized global cognitive factor scores differed by cohort. We first performed a *t*-test to compare the means of the two cohorts. This was followed by a Bayesian linear model that regressed factor scores on age (centered at age 70), a categorical variable representing educational attainment (reference = 10–12 years), gender (reference = female), and cohort (reference = NACC)—plus all interactions—to examine how cognition varied as a function of these predictors and whether the associations between these predictors and cognition differed by cohort. We accounted for measurement error in the estimated factor scores by incorporating individual standard error estimates in brms version 2.19.0.[46]

TABLE 1 Participant demographics.

Characteristic	NACC	VIP
N	15,923	548
Age, mean (SD) [Range]	70.49 (8.42) [50, 103]	72.78 (5.33) [64, 93]
Female gender, *n* (%)	9091 (57.1%)	301 (54.9%)
Education, *n* (%)		
Did not go to school	1 (0.0%)	8 (1.5%)
Primary school (grades 1–5)	15 (0.1%)	72 (13.2%)
Middle school (6–9)	92 (0.6%)	74 (13.6%)
High school (10–12)	2234 (14.1%)	191 (35.0%)
Some college/associate degree	2814 (17.8%)	129 (23.6%)
College/graduate degree	10,675 (67.4%)	72 (13.2%)
Hispanic ethnicity, *n* (%)	1018 (6.4%)	0 (0%)
Race, *n* (%)		
American Indian or Alaska Native	158 (1.0%)	0 (0%)
Asian	438 (2.8%)	548 (100%)
Black or African American	2733 (17.2%)	0 (0%)
Native Hawaiian or Other Pacific Islander	20 (0.1%)	0 (0%)
White	12,377 (77.7%)	0 (0%)
Other	116 (0.7%)	0 (0%)
Unknown	81 (0.5%)	0 (0%)

Abbreviations: NACC, National Alzheimer's Coordinating Center; SD, standard deviation; VIP, Vietnamese Insights into Cognitive Aging Program.

3 | RESULTS

Participant descriptive characteristics, broken down by cohort, are shown in Table 1.

3.1 | Item banking

Models were first established in separate groups to ensure that the specified models fit the data well in each cohort and to obtain item parameter estimates for banking. Results showed that the models fit well in the VIP cohort (C2 ($df = 133$) = 180.50, $p = 0.004$; CFI = 0.993;

TLI = 0.991; RMSEA = 0.026, 90% "confidence" interval [CI]: 0.015 to 0.035; SRMR = 0.037) and the NACC cohort (C2 ($df = 117$) = 4344.61, $p < 0.001$; CFI = 0.979; TLI = 0.972; RMSEA = 0.052, 90% CI: 0.051 to 0.054; SRMR = 0.107). Parameter estimates from these models were saved in item banks for each cohort.

3.2 | DIF

Next, we sought to determine whether any of the potential linking items contained DIF. The DIF detection procedure identified five of the seven linking items as having DIF across the two cohorts. The two items without DIF according to the SABIC statistic were Animal Fluency and Number Span Backward. Results of the DIF detection procedures are provided in Table S4. Cohort-specific parameter estimates for the five items flagged as having DIF are shown in Table S5.

With pooling across the DIF-adjusted and unadjusted IRT models, participants' factor scores were estimated with an average standard error of 0.32 standard deviation (SD) units (95% CI: 0.27 to 0.41). Comparison of the DIF-adjusted factor scores to the unadjusted factor scores showed that 0 of the NACC participants' and 12 (2.19%) of the VIP participants' estimated factor scores were influenced by salient DIF. In other words, 12 participants' factor scores differed by a greater margin than their pooled standard error. The average pooled standard error in participants flagged as having salient DIF was 0.31 SD units, compared to 0.35 SD units in the participants without salient DIF. A boxplot depicting the distribution of the scaled difference scores in the VIP cohort is shown in Figure 2. The unadjusted factor scores were very highly correlated with the DIF-adjusted factor scores, both in the entire sample ($r \tilde{\neq} 1.000$) and in the separate cohorts ($r_{NACC} \tilde{\neq} 1.000$; $r_{VIP} = 0.992$).

3.3 | Scoring

Having established that the unadjusted factor scores were unlikely to be influenced by salient DIF, we generated harmonized factor scores using all seven shared items as linking items. The parameter estimates from the unadjusted model are shown in Table 2. A scatterplot depicting the association between factor score estimates and standard error estimates is shown in Figure 3. The x-axis of Figure 3 shows the estimated factor scores (estimated global cognitive ability levels) for each participant. The precision of these estimates, expressed as standard errors, is depicted on the y-axis. The distributional properties of these scores are plotted in the top margin for the factor scores and in the right margin for the standard errors. In general, the factor scores in NACC were slightly higher (top margin) and more precisely estimated (right margin) than in VIP. In NACC, the factor scores were most precisely estimated in the middle of the ability range (roughly −2 to 2 SDs around the mean), with precision decreasing as estimates became more extreme in both directions. VIP factor score estimates tended to be most precise at lower ability levels, with standard errors increasing relatively linearly as ability level increased.

3.4 | Group comparisons

Factor score means for VIP and NACC were compared using a Bayesian independent samples *t*-test that did not assume equal variances. Results showed that the mean factor score in VIP was ≈0.30 (95% CI: 0.25 to 0.36) points lower than the mean of the NACC cohort. However, because of demographic differences between the two cohorts (Table 1), we also sought to understand how cognition differed in these cohorts using a model that accounted for age, education, and gender. Results of a Bayesian model regressing the factor scores on cohort, age, education, gender, and their interactions are shown in Table 3. As expected, older age was associated with lower factor scores; this effect was largely consistent across cohorts. The effect of gender on factor scores differed quite a bit by cohort; in VIP, being male—compared to being female—was associated with better cognition, whereas the opposite pattern was observed in NACC. Finally, the associations between education and cognition showed modest differences between cohorts, but these effects were not statistically distinguishable from 0 with 95% credibility.

We performed a sensitivity analysis by running the same statistical model, but with DIF-adjusted factor scores instead of unadjusted factor scores as the outcome. We sought to determine whether demographic associations with cognition differed depending on whether corrections for DIF were applied. Results, presented in Table S6, show that none of the differences in parameter estimates were outside of those estimates' margin of error.

4 | DISCUSSION

It is essential that researchers are equipped with appropriate tools to minimize bias when measuring cognition in underrepresented populations. The current study was conducted to harmonize a comprehensive measure of global cognition in a novel cohort of older Vietnamese American immigrants on a metric that is scaled to match one of the largest and well-characterized samples of older adults in the world: the NACC UDS.

Three primary findings emerged from this study. The first is that global cognitive ability in VIP can be measured with good precision using a battery of cognitive tests derived from the UDS 3 neuropsychological battery, the CASI, and the WHO-UCLA AVLT.[30] In 98% of the VIP cohort, standard errors were less than 0.4, which corresponds to a reliability (similar to internal consistency) of >0.86. The second is that, even though five of the seven items common to the two cohorts showed evidence of DIF, the impact of this DIF on the resulting factor score estimates was negligible. This means that we were able to use the maximum number of linking items to generate harmonized factor scores, which can help avoid biased harmonization.[33] Finally, the third finding is that the IRT-based harmonization methods allow for valid group mean comparisons to be made between VIP and NACC with confidence that any observed differences are minimally influenced by systematic bias associated with language or cultural factors.

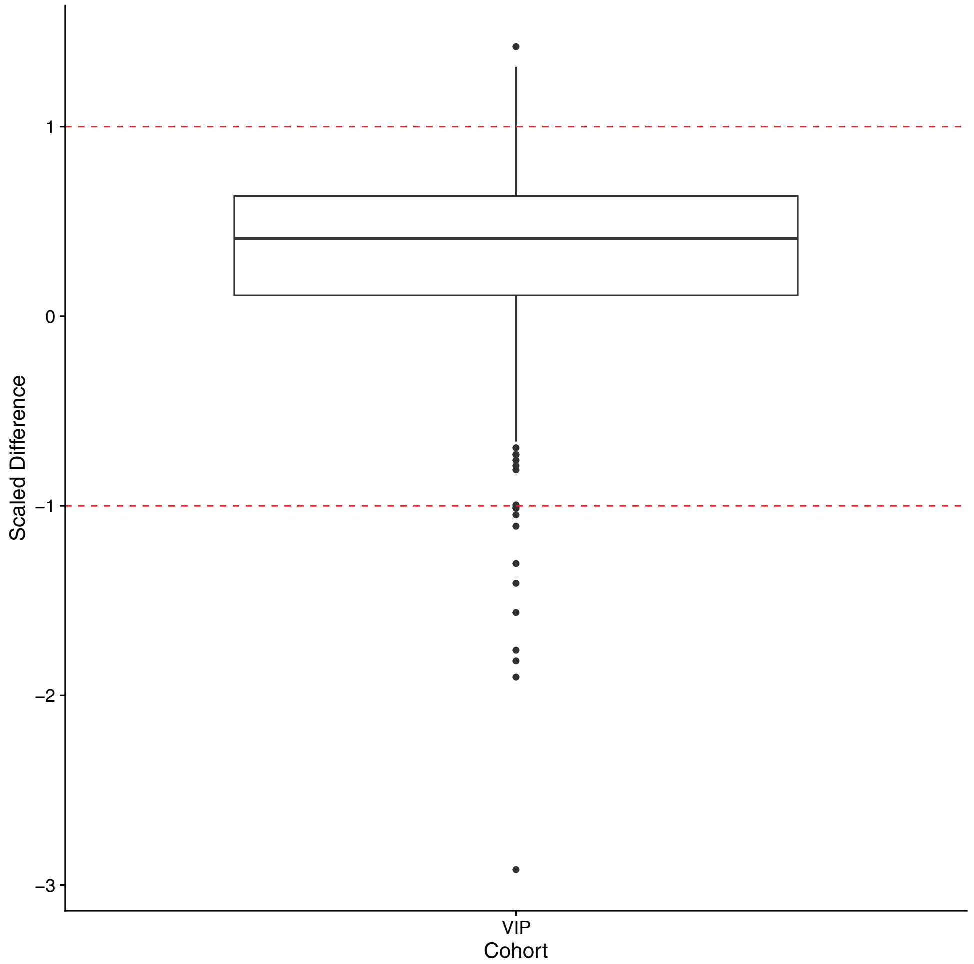

FIGURE 2 Boxplot showing the distribution of the difference in the uncorrected factor score estimates relative to the DIF-corrected factor score estimates, scaled by the pooled standard error. The horizontal dashed red lines depict ± 1 standard error, which was our threshold for detecting salient DIF. DIF, differential item functioning.

The harmonized factor scores that were generated in the current study were used to understand how three demographic variables (age, gender, and education) impact global cognitive performance in the combined sample and whether the strength of association between these variables and global cognition varies by cohort. Only one of these demographic variables—gender—was differentially associated with global cognition across cohorts. In NACC, the average estimated factor score in men was –0.11 points lower than in women when education and age were held constant. In contrast, the average estimated factor score was 0.23 points higher in men compared to women in VIP, also holding education and age constant. Although our results did not find robust evidence to suggest that these gender differences could be explained by differences in education, our findings cannot conclusively reject such a hypothesis. Results in Table 3 show that the three-way interactions between gender × education × cohort yielded some of the largest effect sizes; however, due to the large standard errors estimating these effects, it is impossible to confidently interpret these findings. Given the importance of education and occupational status in promoting late-life cognition,[47–50] future research may wish to further explore cross-national and cross-cultural differences in barriers to accessing education and other cognitively stimulating activities across people of different genders.

The current study represents an important step forward in the science of cognitive aging by helping to reduce barriers to making valid estimates of cognition in an important, underrepresented, and understudied population. Most published research assessing cognition in Vietnamese individuals has relied on basic screening measures, such as the Mini-Mental State Examination (MMSE)[51–53] or the MoCA.[54] Although these measures have been translated from English to Vietnamese, we are unaware of any work that has sought to determine whether the measurement properties of the translated scores are equivalent in both languages and whether any items show evidence of DIF. Thus there is a clear need for more comprehensive and psychometrically sound options for performing cognitive assessment in Vietnamese speakers.

One approach to comprehensive neuropsychological assessment of Vietnamese-speaking individuals was described by McCauley and colleagues, who proposed a battery of assessment instruments selected for use in a clinical setting.[55] McCauley et al. described a non-native English-speaking patient who performed markedly better

TABLE 2 Item parameter estimates used to generate harmonized factor scores (slope-intercept form).

Type	Item	a1	a2	a3	a4	a5	d1	d2	d3	d4	d5	d6	d7	d8	d9
Linking item	Benson Delayed Recall	2.07	1.49				3.88	2.60	1.77	0.71	0.12	−0.45	−1.11	−1.77	−2.50
Linking item	Benson Recognition	1.09	0.62				1.66								
Linking item	Trail Making Test Part A	3.16		2.36			8.18	7.59	6.93	6.08	5.19	4.42	3.27	1.75	−0.43
Linking item	Benson Figure Copy	1.04		0.12			3.76	3.10	2.31	1.67	0.82	−0.36	−1.76		
Linking item	Number Span Forward	1.71			2.51		4.42	2.89	1.14	−0.12	−1.52	−2.84	−4.17	−5.41	−6.73
Linking item	Number Span Backward	2.05			1.05		6.00	4.14	2.95	1.44	0.44	−0.72	−1.88	−3.09	−4.17
Linking item	Animal Fluency	3.80				2.43	6.44	4.92	3.30	1.81	1.09	−0.36	−1.72	−2.36	−4.26
NACC item	MoCA Word Recall	2.03	1.51				1.43	0.66	−0.37	−1.61	−3.11				
NACC item	MoCA Orientation	2.29	1.28				7.60	6.18	5.15	4.20	3.16	1.46			
NACC item	Craft Story	2.18	1.18				3.82	2.71	1.82	0.71	−0.04	−0.67	−1.14	−2.06	−3.05
NACC item	Trail Making Test Part B	3.65		1.38			3.85	2.12	0.94	−0.04	−0.97	−1.82	−2.76	−3.90	−5.52
NACC item	MoCA Trail Making	1.49		0.27			0.94								
NACC item	MoCA Clock Drawing	1.81		0.09			5.49	2.42	0.12						
NACC item	MoCA Sentence Repetition	1.24			0.72		2.20	0.05							
NACC item	MoCA Serial 7s	2.07			0.26		3.78	2.47	0.89						
NACC item	Vegetable Fluency	2.21				0.95	4.84	3.44	2.19	1.58	1.02	0.46	−0.07	−1.11	−2.65
NACC item	Letter Fluency	1.63				0.53	4.05	2.58	1.54	0.51	0.02	−0.46	−0.91	−1.86	−2.62
NACC item	Multilingual Naming Test	1.56				0.44	3.92	2.79	1.99	1.54	1.03	0.45	−0.23	−1.04	−2.00
VIP item	AVLT Trial 7	1.31	2.56				3.96	2.83	2.13	1.33	0.39	−0.74	−1.92	−3.13	−4.50
VIP item	AVLT Trial 6	1.16	2.51				5.78	3.58	2.42	1.39	0.15	−0.67	−1.85	−2.88	−4.25
VIP item	AVLT Trial 5	1.24	1.82				4.43	3.37	2.03	0.95	−0.18	−1.32	−2.35	−4.01	−5.52
VIP item	CASI Short-Term Memory	0.98	0.59				3.41	2.43	1.70	1.04	0.27	−0.40	−1.16	−2.08	−3.53
VIP item	CASI Orientation	0.87	0.05				4.04	3.14	2.43	0.88					
VIP item	CASI Long-Term Memory	1.45	−0.16				4.77	1.78							
VIP item	WAIS-IV Coding	2.90		2.26			5.11	3.17	1.53	0.64	−0.24	−1.29	−2.24	−3.50	−5.05
VIP item	CASI Drawing	1.19		0.42			4.77	3.72	3.16	1.97	0.94				
VIP item	CASI Mental Manipulation	1.57			1.18		5.60	4.35	3.60	2.53	1.00	0.40	−1.31	−1.80	
VIP item	CASI Attention	1.50			0.04		4.71	3.30	2.16	1.00	−0.65				
VIP item	CASI Language	1.49				0.75	4.46	3.52	1.80						
VIP item	CASI Abstraction and Judgment	1.63				0.71	5.01	3.61	1.48	−0.02	−1.22	−1.97	−3.22	−4.48	

Abbreviations: a1, item slope parameter for global factor; a2, item slope parameter for memory factor; a3, item slope parameter for visual attention and executive functioning factor; a4, item slope parameter for auditory attention and working memory factor; a5, item slope parameter for language factor; AVLT, Auditory Verbal Learning Test; CASI, Cognitive Abilities Screening Instrument; d1-d9, item intercept parameters; MoCA, Montreal Cognitive Assessment; NACC, National Alzheimer's Coordinating Center; VIP, Vietnamese Insights into Cognitive Aging Program; WAIS-IV, Wechsler Adult Intelligence Scales, 4th Edition.

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

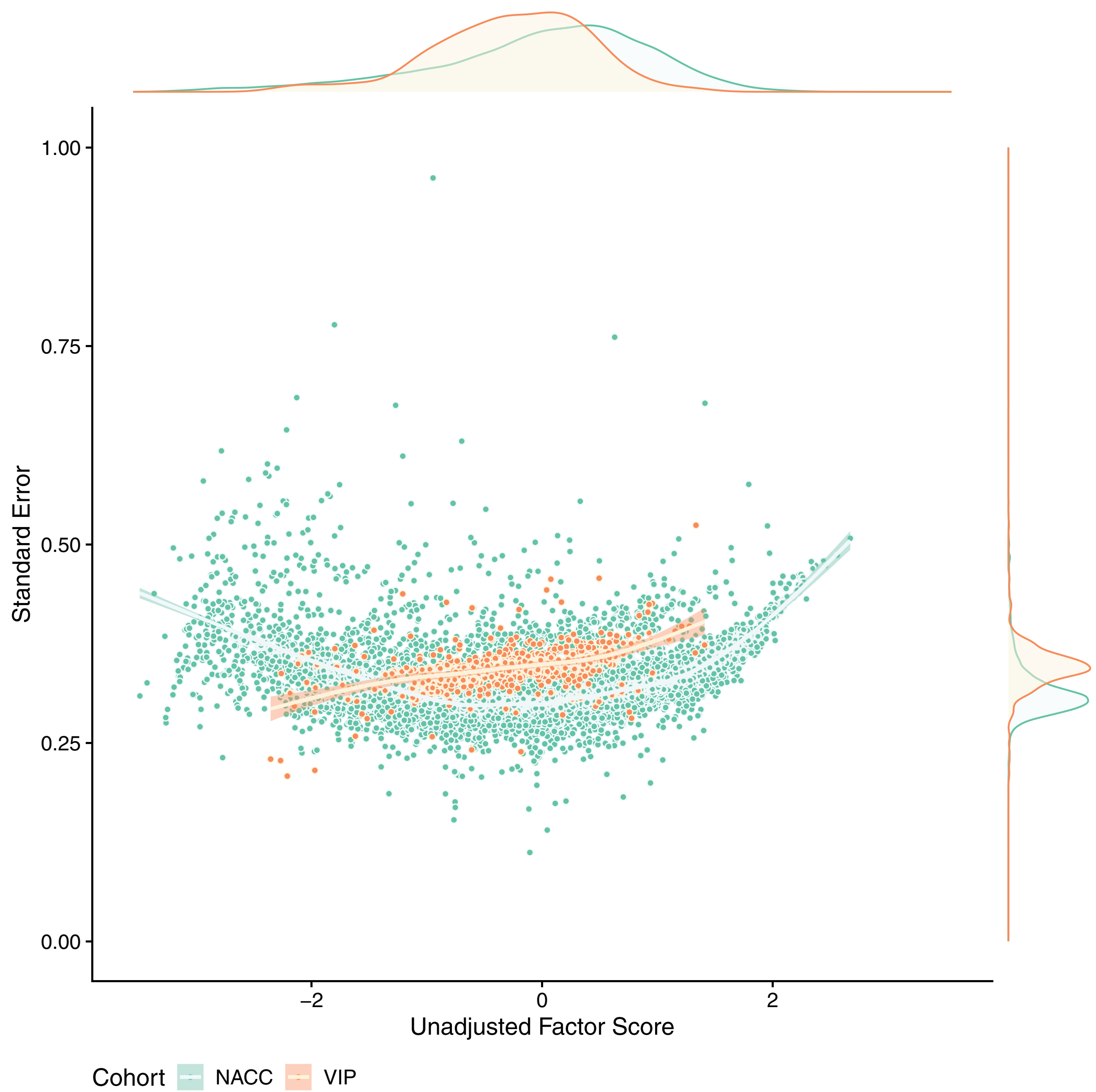

FIGURE 3 Scatterplot showing the associations between the unadjusted factor score estimates (x-axis) and the standard errors of these estimates (y-axis) in the NACC and VIP cohorts. Density curves depict the distributional shapes of the factor scores (top margin) and standard errors (right margin). NACC, National Alzheimer's Coordinating Center; VIP, Vietnamese Insights into Cognitive Aging Program.

when assessed in Vietnamese ≈1 year after having been administered the same battery in English.[55] This improvement in test scores might have occurred due to greater proficiency in Vietnamese than in English or may be a function of factors such as practice effects and measurement non-invariance (or DIF). This highlights the importance of evaluating and—if necessary—accounting for DIF and the importance of ensuring measurements produce outcomes that are on the same numeric scale, as the absence of (or adjustment for) DIF can facilitate the identification of other variables that contribute to differences in latent cognitive ability.

In the current study, five of seven items were found to have evidence of DIF: Benson Figure Copy, Benson Delayed Recall, Benson Recognition, Number Span Forward, and Trail Making Test part A (see Tables S4 and S5). Although the combined impact of this DIF on the global factor score estimates was negligible, it is important to note that some DIF occurred, even though we were conscientious about the selection and adaptation of these items with considerations for cultural and linguistic fairness.[30] Furthermore, a team of bilingual and bicultural researchers and staff (some of whom were not part of the original test adaptation process) rated the items as "essentially equivalent" after having

TABLE 3 Coefficients for the regression of the harmonized global cognitive factor score on demographics.

Term	Estimate	SE	Lo95	Hi95
(Intercept)	−0.464[a]	0.023	−0.510	−0.417
Age (centered at age 70)	−0.01[a]	0.003	−0.015	−0.005
Male gender	−0.112[a]	0.038	−0.187	−0.04
Education: < 10 years	−0.188	0.116	−0.412	0.04
Education: Some college/associate degree	0.366[a]	0.031	0.306	0.426
Education: College/graduate degree	0.729[a]	0.026	0.679	0.78
VIP cohort	0.250[a]	0.097	0.060	0.441
Age × Male	0.010[a]	0.004	0.000	0.017
Age × < 10 years edu	−0.012	0.010	−0.032	0.008
Age × Some/Associate edu	−0.003	0.004	−0.010	0.004
Age × College/Graduate edu	−0.006[a]	0.003	−0.012	−0.000
Male × < 10 years edu	0.046	0.184	−0.310	0.401
Male × Some/Associate edu	−0.026	0.051	−0.124	0.072
Male × College/Graduate edu	−0.099[a]	0.041	−0.179	−0.019
Age × VIP	−0.029	0.018	−0.066	0.006
Male × IP	0.338[a]	0.149	0.048	0.630
VIP × < 10 years edu	−0.334	0.175	−0.685	0.004
VIP × Some/Associate edu	−0.040	0.152	−0.332	0.261
VIP × College/Graduate edu	−0.310	0.204	−0.716	0.081
Age × Male × < 10 years edu	−0.007	0.016	−0.040	0.024
Age × Male × Some/Associate edu	−0.003	0.006	−0.015	0.008
Age × Male × College/Graduate edu	−0.005	0.005	−0.015	0.004
Age × Male × VIP cohort	−0.010	0.025	−0.060	0.041
Age × VIP × < 10 years edu	0.010	0.025	−0.038	0.059
Age × VIP × Some/Associate edu	0.014	0.031	−0.046	0.075
Age × VIP × College/Graduate edu	0.024	0.043	−0.060	0.110
Male × VIP × < 10 years edu	−0.158	0.297	−0.725	0.439
Male × VIP × Some/Associate edu	−0.255	0.246	−0.737	0.228
Male × VIP × College/Graduate edu	0.120	0.283	−0.422	0.669
Age × Male × VIP × < 10 years edu	0.031	0.040	−0.048	0.110
Age × Male × VIP × Some/Associate edu	0.003	0.044	−0.082	0.088
Age × Male × VIP × College/Graduate edu	−0.021	0.053	−0.126	0.083

Note: The reference category for Education is 10–12 years.
Abbreviations: edu, education; Hi95, upper boundary of the 95% credible intervals; Lo95, lower boundary of the 95% credible intervals; SE, standard error; VIP, Vietnamese Insights into Cognitive Aging Program.
[a]95% Credible intervals do not include 0.

spent more than 2 years administering and interpreting the resulting test scores. Thus, it is important to emphasize that expert opinion and clinical experience do not necessarily match empirical findings; it is always important to rigorously test hypotheses about item-level and scale-level equivalence.[27] This finding also highlights the limitations of modifying existing test instruments for use in other language and/or cultural groups, rather than prospectively developing tests with broad applicability in mind at the outset.

Despite its strengths, the current study is not without limitations. One of the most salient limitations is that, due to the design of the cognitive battery and the limited number of items shared between VIP and NACC, we were only able to harmonize the global cognition factor across cohorts. Although our model estimated factors for memory, auditory attention and working memory, visual attention and executive functions, and language, these specific factors were used to account for domain-specific variance that would not have been accounted for in a uni-dimensional model. It may have been possible to attempt harmonization in each of the four specific cognitive domains, but only one to two items were shared across cohorts, increasing the risk of biased harmonization.[33] Nevertheless, such an investigation could be pursued in future research. Global cognitive scores can be useful in many scenarios but lack specificity for other applications.[56] That the harmonization was limited to a global composite score imposes limits on the comparisons that can be made across cohorts without precluding within-cohort comparisons using non-harmonized measures. An additional limitation of the current study is that our focus was on making comparisons across VIP and NACC cohorts. However, other sources of individual differences are inherent in these cohorts, such as sex/gender, age, education, and region of birth/upbringing (e.g., North vs South may be relevant in both Vietnam and the United States). We did not attempt to investigate the presence of DIF as a function of any of these other characteristics. Similarly, the current study did not attempt to understand whether the global factor scores were invariant to time, but this will certainly be a target of future research.

In conclusion, the current study establishes that global cognitive functioning can be estimated with minimal bias in a cohort of Vietnamese American immigrants in a way that is psychometrically matched to one of the largest studies of cognitive aging and dementia worldwide. The large sample size and clinical heterogeneity in NACC ensures excellent coverage of the full range of ability levels measured by these neuropsychological tests and precise estimation of parameters used as linking items. These findings provide a mechanism by which cognitive performance can be validly assessed in a highly underrepresented group in cognitive aging research, creating possibilities to better understand in future research how factors such as war-related trauma, migration, and asylum-seeking contribute to cognitive aging.

ACKNOWLEDGMENTS

The authors would like to thank the community advisory boards, two community partners involved in the Vietnamese Insights into Cognitive Aging Program (VIP)—Asian Resources, Inc. (ARI) and International Children Assistance Network (ICAN), and the supporting staff. We also want to thank Dr. Malcolm Dick for his guidance and input on the topic of neuropsychological testing in Vietnamese Americans. A copy of the pre-statistical harmonization survey used in this study can be obtained by emailing the corresponding author. The NACC database is funded by National Institute on Aging (NIA)/National Institutes of Health (NIH) Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), and P30 AG072959 (PI James Leverenz, MD). This work is partially supported by the National Institute on Aging [all authors: R01AG067541; O.M., A.K., L.H., Q.T., Q.V., and V.T.P.: R24AG063718]. O.M., S.F., D.H., and R.W. were supported by P30AG072972.

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest. The funders had no role in study conception, design, or writing of this manuscript. Author disclosures are available in the Supporting Information.

CONSENT STATEMENT

All human subjects in the VIP study provided informed consent consistent with the Declaration of Helsinki. Data obtained from NACC were de-identified, but consent was received at the local NACC sites during original data collection, as overseen by each institution's own ethics board.

ORCID

Brandon E. Gavett https://orcid.org/0000-0003-1938-1854

REFERENCES

1. Nápoles AM, Chadiha LA. Advancing the science of recruitment and retention of ethnically diverse populations. *Gerontologist*. 2011;51:S142-146. doi:10.1093/geront/gnr019
2. Maestre G, Hill C, Griffin P, et al. Promoting diverse perspectives: addressing health disparities related to Alzheimer's and all dementias. *Alzheimers Dement*. 2024;20:3099-3107. doi:10.1002/alz.13752
3. Chin AL, Negash S, Hamilton R. Diversity and disparity in dementia: the impact of ethnoracial differences in Alzheimer disease. *Alzheimer Dis Assoc Disord*. 2011;25:187-195. doi:10.1097/WAD.0b013e318211c6c9
4. Peterson RL, George KM, Gilsanz P, et al. Racial/ethnic disparities in young adulthood and midlife cardiovascular risk factors and late-life cognitive domains: the Kaiser Healthy Aging and Diverse Life Experiences (KHANDLE) study. *Alzheimer Dis Assoc Disord*. 2021;35:99-105. doi:10.1097/WAD.0000000000000436
5. Tsoy E, Kiekhofer RE, Guterman EL, et al. Assessment of racial/ethnic disparities in timeliness and comprehensiveness of dementia diagnosis in California. *JAMA Neurol*. 2021;78:657-665. doi:10.1001/jamaneurol.2021.0399
6. Zhu CW, Gu Y, Cosentino S, Kociolek AJ, Hernandez M, Stern Y. Racial/ethnic disparities in misidentification of dementia in medicare claims: results from the Washington Heights-Inwood Columbia Aging project. *J Alzheimers Dis*. 2023;96:359-368. doi:10.3233/JAD-230584
7. Hinton L, Tran D, Peak K, Meyer OL, Quiñones AR. Mapping racial and ethnic healthcare disparities for persons living with dementia: a scoping review. *Alzheimers Dement*. 2024;20(4):3000-3020. doi:10.1002/alz.13612
8. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia incidence between six racial and ethnic groups over

14 years. *Alzheimers Dement*. 2016;12(3):216-224. doi:10.1016/j.jalz.2015.12.007

9. Boone KB, Victor TL, Wen J, Razani J, Pontón M. The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Arch Clin Neuropsychol*. 2007;22:355-365. doi:10.1016/j.acn.2007.01.010
10. Loewenstein DA, Argüelles T, Argüelles S, Linn-Fuentes P. Potential cultural bias in the neuropsychological assessment of the older adult. *J Clin Exp Neuropsychol*. 1994;16:623-629. doi:10.1080/01688639408402673
11. Pedraza O, Mungas D. Measurement in cross-cultural neuropsychology. *Neuropsychol Rev*. 2008;18:184-193. doi:10.1007/s11065-008-9067-9
12. Lipnicki DM, Crawford JD, Dutta R, et al. Age-related cognitive decline and associations with sex, education and apolipoprotein E genotype across ethnocultural groups and geographic regions: a collaborative cohort study. *PLoS Med*. 2017;14:e1002261. doi:10.1371/journal.pmed.1002261
13. Kim G, Chiriboga DA, Jang Y, Lee S, Huang C-H, Parmelee P. Health status of older Asian Americans in California. *J Am Geriatr Soc*. 2010;58:2003-2008. doi:10.1111/j.1532-5415.2010.03034.x
14. Ro A, Geronimus A, Bound J, Griffith D, Gee G. Educational gradients in five Asian immigrant populations: do country of origin, duration and generational status moderate the education-health relationship?. *Prev Med Rep*. 2016;4:338-343. doi:10.1016/j.pmedr.2016.07.001
15. Meyer OL, Park VT, Kanaya AM, et al. Inclusion of Vietnamese Americans: opportunities to understand dementia disparities. *Alzheimers Dement*. 2023;9:e12392. doi:10.1002/trc2.12392
16. Bürgin D, Anagnostopoulos D, Board and Policy Division of ESCAP, et al. Impact of war and forced displacement on children's mental health-multilevel, needs-oriented, and trauma-informed approaches. *Eur Child Adolesc Psychiatry*. 2022;31:845-853. doi:10.1007/s00787-022-01974-z
17. Teicher MH. Childhood trauma and the enduring consequences of forcibly separating children from parents at the United States border. *BMC Medicine*. 2018;16:146. doi:10.1186/s12916-018-1147-y
18. Mungas D, Widaman KF, Reed BR, Tomaszewski Farias S. Measurement invariance of neuropsychological tests in diverse older persons. *Neuropsychology*. 2011;25:260-269. doi:10.1037/a0021090
19. Teresi JA, Holmes D. Some methodological guidelines for cross-cultural comparisons. In: Skinner JH, Teresi JA, Holmes D, Stahl SM, Stewart AL, eds. *Multicultural measurement in older populations*. Springer Publishing Company, Inc.; 2002:3-10.
20. Prince M. Measurement validity in cross-cultural comparative research. *Epidemiol Psychiatric Sci*. 2008;17:211-220. doi:10.1017/S1121189X00001305
21. Jones RN. Differential item functioning and its relevance to epidemiology. *Curr Epidemiol Rep*. 2019;6:174-183. doi:10.1007/s40471-019-00194-5
22. Chan KS, Gross AL, Pezzin LE, Brandt J, Kasper JD. Harmonizing measures of cognitive performance across international surveys of aging using item response theory. *J Aging Health*. 2015;27:1392-1414. doi:10.1177/0898264315583054
23. Meyer OL, Tomaszewski Farias S, Whitmer RA, et al. Vietnamese Insights Into Cognitive Aging Program (VIP): objectives, study design and cohort description. *Alzheimers Dement*. 2024;10:e12494.
24. Arce Rentería M, Briceño EM, Chen D, et al. Memory and language cognitive data harmonization across the United States and Mexico. *Alzheimers Dement*. 2023;15:e12478. doi:10.1002/dad2.12478
25. Briceño EM, Arce Rentería M, Gross AL, et al. A cultural neuropsychological approach to harmonization of cognitive data across culturally and linguistically diverse older adult populations. *Neuropsychology*. 2023;37:247-257. doi:10.1037/neu0000816
26. Gross AL, Li C, Briceño EM, et al. Harmonisation of later-life cognitive function across national contexts: results from the harmonized cognitive assessment protocols. *Lancet Healthy Longev*. 2023;4:e573-583. doi:10.1016/S2666-7568(23)00170-8
27. Vonk JMJ, Gross AL, Zammit AR, et al. Cross-national harmonization of cognitive measures across HRS HCAP (USA) and LASI-DAD (India). *PLoS One*. 2022;17:e0264166. doi:10.1371/journal.pone.0264166
28. Besser L, Kukull W, Knopman DS, et al. Version 3 of the National Alzheimer's Coordinating Center's uniform data set. *Alzheimers Dement*. 2018;32:351. doi:10.1097/WAD.0000000000000279
29. Weintraub S, Besser L, Dodge HH, et al. Version 3 of the Alzheimer Disease Centers' neuropsychological test battery in the Uniform Data Set (UDS). *Alzheimer Dis Assoc Disord*. 2018;32:10-17. doi:10.1097/WAD.0000000000000223
30. Tiet QQ et al. The Translation with Ongoing Adaptation and Improvement (ToAI) framework: Tailoring neuropsychological assessment instruments for the Vietnamese Insights into Cognitive Aging Program (VIP) under review.
31. Teng EL, Hasegawa K, Homma A, et al. The Cognitive Abilities Screening Instrument (CASI): a practical test for cross-cultural epidemiological studies of dementia. *Int Psychogeriatr*. 1994;6:45-58. doi:10.1017/S1041610294001602
32. Maj M, D'Elia L, Satz P, et al. Evaluation of two new neuropsychological tests designed to minimize cultural bias in the assessment of HIV-1 seropositive persons: a WHO study. *Arch Clin Neuropsychol*. 1993;8:123-135.
33. Gavett BE, Ilango SD, Koscik R, et al. Harmonization of cognitive screening tools for dementia across diverse samples: a simulation study. *Alzheimers Dement*. 2023;15:e12438. doi:10.1002/dad2.12438
34. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing; 2024.
35. Chalmers RP. Mirt: a multidimensional item response theory package for the R environment. *J Stat Softw*. 2012;48:1-29. doi:10.18637/jss.v048.i06
36. Cai L, Monroe SL. A new statistic for evaluating item response theory models for ordinal data. CRESST Report 839. National Center for Research on Evaluation, Standards, and Student Testing; 2014.
37. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1-55. doi:10.1080/10705519909540118
38. Kolen M, Brennan R. *Test equating, scaling, and linking: methods and practices*. 3rd Ed. Springer; 2014.
39. Nugent WR. Understanding DIF and DTF: description, methods, and implications for social work research. *J Soc Social Work Res*. 2017;8:305-334. doi:10.1086/691525
40. Sclove SL. Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*. 1987;52:333-343. doi:10.1007/BF02294360
41. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19:716-723. doi:10.1109/TAC.1974.1100705
42. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461-464. doi:10.1214/aos/1176344136
43. Hannan EJ, Quinn BG. The determination of the order of an autoregression. *J R Stat Soc Series B Stat Methodol*. 1979;41:190-195. doi:10.1111/j.2517-6161.1979.tb01072.x
44. Kopf J, Zeileis A, Strobl C. Anchor selection strategies for DIF analysis. *Educ Psychol Meas*. 2015;75:22-56. doi:10.1177/0013164414529792
45. Embretson SE, Reise SP. *Item response theory for psychologists*. Lawrence Erlbaum Associates Publishers; 2000.
46. Bürkner P-C. brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw*. 2017;80:1-28. doi:10.18637/jss.v080.i01
47. Lamar M, Lerner AJ, James BD, Yu L, Glover CM, Wilson RS, et al. Relationship of early-life residence and educational experience to level and change in cognitive functioning: results of the Minority Aging Research Study. *J Gerontol B Psychol Sci Soc Sci*. 2020;75:e81-92. doi:10.1093/geronb/gbz031

48. Lövdén M, Fratiglioni L, Glymour MM, Lindenberger U, Tucker-Drob EM. Education and cognitive functioning across the life span. *Psychol Sci Public Interest*. 2020;21:6-41. doi:10.1177/1529100620920576
49. Owens JH, Fiala J, Jones RN, Marsiske M. The mediating effects of education and occupational complexity between race and longitudinal change in late life cognition in ACTIVE. *Res Aging*. 2024;46:492-508. doi:10.1177/01640275241248825
50. Ritchie SJ, Tucker-Drob EM. How much does education improve intelligence? A meta-analysis. *Psychol Sci*. 2018;29:1358-1369. doi:10.1177/0956797618774253
51. Korinek K, Zimmer Z, Teerawichitchainan B, Young Y, Cao Manh L, Toan TK. Cognitive function following early life war-time stress exposure in a cohort of Vietnamese older adults. *Soc Sci Med*. 2024;349:116800. doi:10.1016/j.socscimed.2024.116800
52. Miyawaki CE, Garcia JM, Nguyen KN, Park VT, Markides KS. Multiple chronic conditions and disability among Vietnamese older adults: results from the Vietnamese Aging and Care Survey (VACS). *J Rac Ethnic Health Dispar*. 2024;11:1800-1807. doi:10.1007/s40615-023-01652-z
53. Nguyen VT, Quach THT, Pham AG, Tran TC. Feasibility, reliability, and validity of the Vietnamese version of the Clinical Dementia Rating. *Dement Geriatr Cogn Disord*. 2020;48:308-316. doi:10.1159/000506126
54. Nguyen TT-Q, Hoang CBD, Hoang Le MD, et al. Assessing cognitive decline in Vietnamese older adults using the Montreal Cognitive Assessment-Basic (MoCA-B) and Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) during the COVID-19 pandemic: a feasibility study. *Clin Neuropsychol*. 2023;37:1043-1061. doi:10.1080/13854046.2023.2192418
55. McCauley SR, Nguyen T, Nguyen C, Strutt AM, Stinson JM, Windham VA, et al. Developing a culturally competent neuropsychological assessment battery for vietnamese-speaking patients with suspected dementia. *Arch Clin Neuropsychol*. 2023;38:485-500. doi:10.1093/arclin/acac035
56. Mukherjee S, Choi SE, Lee ML, et al. Cognitive domain harmonization and cocalibration in studies of older adults. *Neuropsychology*. 2023;37(4):409-423. doi:10.1037/neu0000835

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gavett BE, Tomaszewski Farias S, Tiet QQ, et al. Harmonized cognitive performance in an older adult cohort of Vietnamese American immigrants: The VIP study. *Alzheimer's Dement*. 2025;21:e70097. https://doi.org/10.1002/alz.70097