Journal of Gerontology: PSYCHOLOGICAL SCIENCES
1998, Vol. 53B, No. 6, P370–P374

Copyright 1998 by The Gerontological Society of America

An Analysis of Test Bias and Differential Item Functioning Due to Race on the Mattis Dementia Rating Scale

John L. Woodard,[1] Alexander P. Auchus,[1] Robert E. Godsall,[1] and Robert C. Green[2]

[1]Department of Neurology, Emory University School of Medicine.
[2]Memory Assessment Clinic and Alzheimer's Disease Program, College of Health and Human Sciences, Georgia State University.

The Mattis Dementia Rating Scale (MDRS) is a commonly used cognitive measure designed to assess the course of decline in progressive dementias. However, little information is available about possible systematic racial bias on the items presented in this test. We investigated race as a potential source of test bias and differential item functioning in 40 pairs of African American and Caucasian dementia patients (N = 80), matched on age, education, and gender. Principal component analysis revealed similar patterns and magnitudes across component loadings for each racial group, indicating no clear evidence of test bias on account of race. Results of an item analysis of the MDRS revealed differential item functioning across groups on only 4 of 36 items, which may potentially be dropped to produce a modified MDRS that may be less sensitive to cultural factors. Given the absence of test bias because of race, the observed racial differences on the total MDRS score are most likely associated with group differences in dementia severity. We conclude that the MDRS shows no appreciable evidence of test bias and minimal differential item functioning (item bias) because of race, suggesting that the MDRS may be used in both African American and Caucasian dementia patients to assess dementia severity.

BRIEF cognitive rating scales are commonly used in the clinical diagnosis and long-term management of dementing disorders, and they have become essential tools for disease identification in population-based epidemiological research and for efficacy measurement in clinical trials of new medications to improve cognition in patients with these disorders. The utility of these scales in populations with a mixture of demographic variables assumes that there will be no systematic bias, but this assumption has been increasingly challenged in recent years, particularly with regard to race (Dollear et al., 1994; Fillenbaum, Heyman, Williams, & Burchett, 1990; Gurland, Wilder, Cross, Teresi, & Barrett, 1992; Helms, 1992; Loewenstein, Argüelles, Argüelles, & Linn-Fuentes, 1994; Loewenstein, Argüelles, Barker, & Duara, 1993). For example, epidemiological studies have reported diminished specificity associated with a higher false positive rate (Fillenbaum et al., 1990; Gurland et al., 1992) and a greater tendency for African Americans to be diagnosed as cognitively impaired (Callahan, Hendrie, & Tierney, 1995; Cohen & Carlin, 1993) when they were administered screening instruments such as the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) and the Short Portable Mental Status Questionnaire (Pfeiffer, 1975). These potential sources of variance are of particular importance in the assessment of elderly patients, when judgments about competence and ability to independently execute activities of daily living (ADLs) are often a focal point of the evaluation.

The Mattis Dementia Rating Scale (MDRS; Mattis, 1973) is a popular instrument for assessing and tracking cognitive changes in dementia patients, and portions of the MDRS have been used to screen for cognitive impairment (Green, Woodard, & Green, 1995; Shay et al., 1991; Vitaliano et al., 1984). The test takes approximately 20–45 minutes to administer (Vitaliano et al., 1984), and reliability is high (Smith et al., 1994; van Belle, Uhlmann, Hughes, & Larson, 1990; Vitaliano et al., 1984). When combined with scores from the Instrumental Activities of Daily Living scale, patient assessments have been reported to closely approximate clinical judgments of dementia severity based upon full clinical evaluation (Shay et al., 1991). The MDRS has been shown to be a clinically valid psychometric test for the detection of Alzheimer's disease in a community-based sample (Monsch et al., 1995), and MDRS total score has been shown to be a significant predictor of longitudinal institutionalization and mortality outcomes (Smith et al., 1994) in patients with dementias of various etiologies. However, only two studies, to our knowledge, have reported the influence of race on MDRS performance. In the first study (Vangel & Lichtenberg, 1995), point-biserial correlations were computed between MDRS total score and gender and race, while Pearson product-moment correlations were calculated between MDRS total score and age and education in a sample of 90 cognitively intact individuals. This study reported significant correlations between MDRS total score and age and education, but MDRS total score was not significantly correlated with race or gender.

A second study (Lichtenberg, Ross, Millis, & Manning, 1995) used multiple regression analysis to explore the influence of Geriatric Depression Scale score and demo-

Downloaded from https://academic.oup.com/psychsocgerontology/article/53B/6/P370/618616 by guest on 21 August 2022

graphic variables (age, education, race, and gender) on MDRS total score in derivation and normal cross-validation samples of urban geriatric inpatients hospitalized for orthopedic injuries or stroke. In this study, multiple regression analysis revealed a significant effect for race after controlling for age, education, gender, and Geriatric Depression Scale score, although the nature of the race effect was not a focus of the study and was not directly addressed. Given its frequent clinical use, it is noteworthy that there has been little systematic investigation of possible test bias or patterns of differential item functioning (DIF) attributable to race for the MDRS in dementia patients. This information would be very useful for ensuring the appropriate use of the MDRS with racially heterogeneous populations.

The purpose of this study was to examine possible test bias and DIF due to race on the MDRS in African American and Caucasian dementia patients who were matched on age, education, and gender. Matching on these demographic variables permitted us to examine possible test bias due to race independent of possible preexisting group differences on age and education, and severity of dementia. Techniques used in this analysis consisted of: (a) reliability (internal consistency) analysis, (b) principal components analysis, and (c) a combination of multiple and logistic regression procedures for each of the MDRS items to detect DIF, thereby revealing those items that might show a different pattern of performance for each racial group (Ashford, Kolm, Colliver, Bekian, & Hsu, 1989; Swaminathan & Rogers, 1990).

Method

Participants

Participants consisted of a convenience sample of individuals seen at two memory assessment clinics, one serving predominately low-income African American patients, and one serving mostly middle- and upper-income Caucasian patients. From this overall sample, African American dementia patients were matched with Caucasian dementia patients on the critical variables of gender, education, and age. This algorithm allowed us to identify 40 pairs of subjects with complete MDRS data. Each group contained 8 men (20%), and both groups were statistically matched ($p > .05$) on age (Caucasian mean = 75.1 years, SD = 6.3; African American mean = 74.2 years, SD = 6.0) and education (Caucasian mean = 8.7 years, SD = 3.2; African American mean = 8.9 years, SD = 3.2). The overall sample included several different dementia etiologies, although the largest single group (43.8%) was diagnosed with probable or possible AD according to NINCDS-ADRDA criteria (McKhann et al., 1984). Diagnoses were made by a board-certified neurologist or licensed clinical psychologist. The African American group had a greater representation of patients with vascular dementia (n = 11) than did the Caucasian group (n = 2) according to NINCDS-AIREN criteria (Erkinjuntti, 1994; Lopez et al., 1994; Román et al., 1993).

These samples were not intended to be representative of the types or severity of dementia, disease duration, age, or education in each population, and it was expected that varying dementia severity and/or disease duration would be present in each group. We focused on studying only patients with a diagnosis of dementia who were matched on age, education, and gender, irrespective of disease duration, severity, or dementia type. In addition, because the MDRS was the only common measure of severity, attempting to match on this variable would have restricted the range of scores and the number of participants available to include in this analysis.

Study Instruments and Examination Procedures

The MDRS consists of 36 items that have been grouped into five content-specific domains: attention, initiation and perseveration, construction, conceptualization, and memory. The measure has been reported to show good reliability (van Belle et al., 1990; Vitaliano et al., 1984), and factor analytic studies have supported the validity of the content-specific domains (Colantonio, Becker, & Huff, 1993; Woodard, Salthouse, Godsall, & Green, 1996). Strong correlations between scores in these content domains and external neuropsychological measures assessing similar cognitive constructs have been reported (Woodard et al., 1996). All participants received the standard administration of the MDRS by a trained psychometrist or licensed psychologist as part of a more comprehensive neuropsychological evaluation, and all examiners were Caucasian.

Data Analysis

We first computed coefficient alpha reliabilities separately for each group in order to determine whether differential test reliability might exist for the two groups. We next tested the hypothesis that the intergroup performance disparity results from bias in the underlying construct(s) measured by the MDRS. Test bias is typically defined with respect to systematic group differences in regression slopes, intercepts, or standard errors of estimate of the regression lines for the two groups when using the test in question to predict an external criterion (Jensen, 1980). However, predictive bias can also be investigated by testing for group differences in pattern and magnitude of loadings obtained from principal components analysis (PCA) or factor analysis (Humphreys & Taber, 1973; Jensen, 1980). The logic behind this approach is that if the relationships between the subtests and the factor underlying performance on a composite measure possess a similar pattern and are of equivalent magnitude across two groups, it is reasonable to expect comparable regression lines for the two groups (Humphreys & Taber, 1973). Thus, a PCA was performed separately for each group, and the resultant component loadings were first subjected to a Fisher's r to z transformation, followed by a chi-square goodness-of-fit analysis to test the null hypothesis that the pattern and magnitude of component loadings were equivalent between groups (Humphreys & Taber, 1973; Jensen, 1980).

We followed these analyses with a series of item characteristic curve analyses using a logistic and multiple regression approach (Swaminathan & Rogers, 1990). For dichotomously scored items, a separate logistic regression analysis was performed for each item in which item response (pass or fail) was predicted by group membership (African American or Caucasian), dementia severity (MDRS total score), and the interaction between group and dementia severity.

Downloaded from https://academic.oup.com/psychsocgerontology/article/53B/6/P370/618616 by guest on 21 August 2022

For items that were not dichotomously scored, a multiple regression approach analogous to the logistic regression approach described above was used. That is, an item's score served as the dependent variable, and group membership, MDRS total score, and the interaction between Group and MDRS total score were the independent variables.

All items should show a significant relationship between MDRS total score and item response if they contribute to estimating the construct assessed by the MDRS. If group membership is significantly related to item response, the item in question is said to exhibit uniform DIF. That is, the probability of a correct response differs significantly between the groups such that the item is consistently easier for one group across all ability levels. Non-uniform DIF is present in an item for which the interaction term is significantly related to item response. Non-uniform DIF suggests that one group will perform better on an item at low ability levels, whereas the item will be harder for that group at high ability levels.

A two-step approach was used to identify DIF (Camilli & Shepard, 1994). First, each item was subjected to either logistic regression (dichotomous items) or multiple regression (non-dichotomous items), entering group (African American or Caucasian), MDRS total score, and the Group × MDRS total score interaction as predictors of the item score. Items for which there was no relationship with MDRS total score and items for which there was a significant ($p < .10$) effect for either group or the Group × MDRS total score interaction were eliminated, and a modified MDRS score was computed based on the sum of the retained items. In the second step, each item was again subjected to either logistic or multiple regression analysis using the same predictors, except that the modified MDRS score was substituted for the MDRS total score. Because the modified MDRS score is based on items not showing DIF in the first step, it serves as an estimate of ability that eliminates the potential effects of item bias on total score. Items that continue to show a significant ($p < .05$) group or interaction effect in the second step are considered to show DIF.

Results

Table 1 illustrates the MDRS total and subtest performance for the two matched dementia groups. The African American sample demonstrated somewhat greater dementia severity relative to the Caucasian sample as measured by the MDRS total score [$t(78) = 2.4$, $p < .05$]. However, individual subtest performance was not significantly different ($ps > .05$) across groups, with the single exception of the Conceptualization subtest score. The Caucasian group obtained significantly higher scores than the African American group on this subtest [$t(78) = 2.9$, $p < .01$].

Table 1. MDRS Scores of Caucasian and African American Dementia Patients Matched on Age, Education, and Gender

	Caucasian		African American	
MDRS Variable	Mean	*SD*	Mean	*SD*
Total MDRS score*	106.7	20.4	96.7	17.9
MDRS Attention	32.3	3.6	31.2	3.4
MDRS Initiation/Perseveration	26.9	7.6	24.2	7.1
MDRS Construction	4.2	1.8	3.7	1.6
MDRS Conceptualization**	28.3	6.8	23.9	6.8
MDRS Memory	15.0	5.4	13.7	5.3

Note: $n = 40$ per group.
*$p < .05$; **$p < .01$.

Internal consistency reliability of the MDRS was computed separately for each group of dementia patients ($n = 40$ per group) using coefficient alpha. No differences in coefficient alpha were seen for the two groups (African American coefficient alpha = .81, Caucasian coefficient alpha = .86), suggesting no psychometric evidence of test bias.

We performed separate PCAs for each racial group ($n = 40$ per group) using the five MDRS subtests. We hypothesized that if the test were biased, different component structures would emerge for the two groups. Both PCAs yielded a single component on which all five subtests loaded, accounting for 60.6% of the variance for the Caucasian group and 53.8% of the variance for the African American group. The magnitude and pattern of factor loadings for each of the five subtests were statistically compared between the two groups using the procedure described by Jensen (1980, p. 449) and differences were found not to be significant ($p > .05$).

Finally, in the first step involving identification of DIF, 7 of the 36 MDRS items were unrelated to total score (single command, imitation, vowel perseveration, draw circle, draw X, write full name, and read word list aloud four times). A significant group effect was observed for alternate tapping, and significant group and group by ability effects were observed for both double alternating movement items (palm up/palm down and fist clenched/fist extended), counting distraction 2 (point out and count As in a random background of letters), and visual recognition memory. Consequently, these 12 items were eliminated and a new ability score was computed using the remaining items. This new score (modified MDRS) thus contained only those items that were significantly related to the total score without demonstrating DIF. In the second step, logistic or multiple regression was performed again for each item, substituting the modified MDRS for MDRS as the dementia severity measure. Significant group effects were noted for both double alternating movement items (palm up/palm down and fist clenched/fist extended), and significant group and group by ability effects were noted for counting distraction 2 (point out and count As) and for visual recognition memory. Thus, only 4 of the 36 items showed DIF. Inspection of the pattern of responses across groups for each of these 4 items revealed a slight advantage for the African American group relative to the Caucasian group.

Discussion

The present study found that only 4 of the 36 MDRS items showed evidence of DIF after matching study participants on age, education, and gender. These four items (palm up/palm down, fist clenched/fist extended, counting distraction 2 [point out and count As], and visual recognition memory) could potentially be eliminated in order to compute a modified MDRS total score that contains no items with DIF. However, despite the presence of four items with DIF on the MDRS, similar reliabilities and pattern and magnitude of component loadings were observed

Downloaded from https://academic.oup.com/psychsocgerontology/article/53B/6/P370/618616 by guest on 21 August 2022

for the African American and Caucasian dementia patient groups, suggesting no evidence for test bias due to race. Furthermore, although our sample of African American dementia patients demonstrated somewhat greater dementia severity than Caucasian dementia patients (as measured by MDRS total score), DIF due to race did not appear to account for this group difference after controlling for dementia severity in our DIF analyses.

There are numerous reports that cognitive measures used for screening for dementia tend to be affected by cultural factors (Fillenbaum et al., 1990; Ford, Haley, Thrower, West, & Harrell, 1996; Gurland et al., 1992; Mungas, Marshall, Weldon, Haan, & Reed, 1996). That is, cognitive screening measures, such as the Mini-Mental State Examination (Folstein et al., 1975), have often shown evidence of bias in minority and low education patients. Cross-cultural differences have also been reported on measures from the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological battery (Welsh et al., 1995) after controlling for other demographic variables, such as age and education. Such performance differences do emphasize that clinicians must be mindful of cultural factors that may contribute to variability in test scores. However, the results of the present study demonstrated an absence of different patterns and magnitudes of component loadings, together with similar reliabilities for the two groups and few items with DIF, providing preliminary evidence suggesting that the MDRS may be used to validly assess dementia severity in both Caucasian and African American dementia patients. It is also possible that by dropping the four items shown to exhibit DIF, the modified MDRS may be even more stable in the face of cultural differences and may maximize sensitivity to true changes in dementia severity. However, given our limited sample size, absence of control participants, and nonpopulation-based sampling approach, we were unable to test the diagnostic utility of the modified MDRS.

It is important to acknowledge the limitations of this study. First, our findings are tempered by the fact that these groups were drawn from two separate clinics serving individuals of differing income levels and referral patterns, and the samples were not of large size and were nonrepresentative of their respective populations in terms of disease duration, dementia severity, income level, and other demographic characteristics not directly examined in this study. A population-based approach to data collection would be far more compelling in order to demonstrate the cross-cultural applicability of the MDRS. The specific issue of disease duration, while difficult to quantify precisely, should also be addressed in future studies. Second, although the groups were equated in terms of the number of years of education, differences in the quality of education are subtler and more difficult to quantify. Third, there are a variety of methods for studying test bias, some of which use internal criteria (such as those methods used in this study), and some of which use external criteria. Although procedures using internal and external ability criteria have been shown to yield largely similar results (Shepard, Camilli, & Averill, 1981), future studies examining possible racial differences in the relationship between MDRS score and external criteria such as presence/absence of dementia or other dementia screening measures would be helpful in further addressing the issue of racial bias in the MDRS. Although the focus of our study was not on diagnostic accuracy, the differential representation of vascular dementia in the two groups represents a potential source of confounding given that the progression of vascular dementia typically differs from that of other degenerative dementias. Finally, study participants were tested by Caucasian examiners, which could potentially have had a differential effect on performance through examiner bias, although this effect cannot be assessed directly in this study.

In summary, we found no psychometric evidence of test bias and only four items with differential item functioning attributable to race in a convenience sample of Caucasian and African American dementia patients matched on gender, age, and education. Our results provide preliminary evidence that the MDRS may be used to validly assess dementia severity in both African American and Caucasian populations, although level and quality of education and other cultural issues that may affect performance must also be considered in the interpretation of MDRS performance.

ACKNOWLEDGMENTS

This research was supported in part by National Institute on Aging Grants P30 AG10130 (to Emory Alzheimer's Disease Center) and R29 AG13912 (FIRST Grant to John L. Woodard).

The authors thank Dr. Felicia C. Goldstein for her helpful comments on an earlier draft of the article.

Dr. John L. Woodard is now at Georgia State University, College of Health and Human Sciences, and Dr. Robert E. Godsall is now at Shepard Pathways, Atlanta, Georgia.

Address correspondence to Dr. John L. Woodard, Memory Assessment Clinic and Alzheimer's Disease Program, Georgia State University, One Park Place South, Suite 801, Atlanta, GA 30303-3083. E-mail: jlwoodard@gsu.edu

REFERENCES

Ashford, J. W., Kolm, P., Colliver, J. A., Bekian, C., & Hsu, L. (1989). Alzheimer patient evaluation and the Mini-Mental State: Item characteristic curve analysis. *Journal of Gerontology: Psychological Sciences*, 44, P139–P146.

Callahan, C. M., Hendrie, H. C., & Tierney, W. M. (1995). Documentation and evaluation of cognitive impairment in elderly primary care patients. *Annals of Internal Medicine, 122,* 422–429.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Thousand Oaks, CA: Sage.

Cohen, C. I., & Carlin, L. (1993). Racial and social differences in clinical and social variables among patients evaluated in a dementia assessment center. *Journal of the National Medical Association, 85,* 379–384.

Colantonio, A., Becker, J. T., & Huff, F. J. (1993). Factor structure of the Mattis Dementia Rating Scale among patients wth probable Alzheimer's disease. *The Clinical Neuropsychologist, 7,* 313–318.

Dollear, T. J., Gorelick, P. B., Dollear, W. C., Harris, Y., Wilson, R. S., & Freels, S. (1994). Comparison of dementia criteria: Sensitivity and specificity testing among African American patients. *Neuroepidemiology, 13,* 59–63.

Erkinjuntti, T. (1994). Clinical criteria for vascular dementia: The NINDS-AIREN criteria. *Dementia, 5,* 189–192.

Fillenbaum, G., Heyman, A., Williams, K., & Burchett, B. (1990). Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly black and white community residents. *Journal of Clinical Epidemiology, 43,* 651–660.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198.

Ford, G. R., Haley, W. E., Thrower, S. L., West, C. A. C., & Harrell, L. E.

Downloaded from https://academic.oup.com/psychsocgerontology/article/53B/6/P370/618616 by guest on 21 August 2022

(1996). Utility of Mini-Mental State Exam scores in predicting functional impairment among White and African American dementia patients. *Journal of Gerontology: Medical Sciences, 51A,* M185–M188.

Green, R. C., Woodard, J. L., & Green, J. (1995). Validity of the Mattis Dementia Rating Scale for detection of cognitive impairment in the elderly. *Journal of Neuropsychiatry and Clinical Neurosciences, 7,* 357–360.

Gurland, B. J., Wilder, D. E., Cross, P., Teresi, J., & Barrett, V. W. (1992). Screening scales for dementia: Toward reconciliation of conflicting cross-cultural findings. *International Journal of Geriatric Psychiatry, 7,* 105–113.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47,* 1083–1101.

Humphreys, L. G., & Taber, T. (1973). Ability factors as a function of advantaged and disadvantaged groups. *Journal of Educational Measurement, 10,* 107–115.

Jensen, A. R. (1980). *Bias in mental testing.* New York: The Free Press.

Lichtenberg, P. A., Ross, T., Millis, S. R., & Manning, C. A. (1995). The relationship between depression and cognition in older adults: A cross-validation study. *Journal of Gerontology: Psychological Sciences, 50B,* P25–P32.

Loewenstein, D. A., Argüelles, T., Argüelles, S., & Linn-Fuentes, P. (1994). Potential cultural bias in the neuropsychological assessment of the older adult. *Journal of Clinical and Experimental Neuropsychology, 16,* 623–629.

Loewenstein, D. A., Argüelles, T., Barker, W. W., & Duara, R. (1993). A comparative analysis of neurosychological test performance of Spanish-speaking and English-speaking patients with Alzheimer's disease. *Journal of Gerontology: Psychological Sciences, 48,* P142–P149.

Lopez, O. L., Larumbe, M. R., Becker, J. T., Rezek, D., Rosen, J., Klunk, W., & DeKosky, S. T. (1994). Reliability of NINDS-AIREN clinical criteria for the diagnosis of vascular dementia. *Neurology, 44,* 1240–1245.

Mattis, S. (1973). *Dementia Rating Scale professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

McKhann, G. M., Drachman, D., Folstein, M. F., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group. *Neurology, 34,* 939–944.

Monsch, A. U., Bondi, M. W., Salmon, D. P., Butters, N., Thal, L. J., Hansen, L. A., Wiederhold, W. C., Cahn, D. A., & Klauber, M. R. (1995). Clinical validity of the Mattis Dementia Rating Scale in detection dementia of the Alzheimer type. *Archives of Neurology, 52,* 899–904.

Mungas, D., Marshall, S., Weldon, M., Haan, M., & Reed, B. (1996). Age and education correction of Mini-Mental State Examination for English and Spanish-speaking elderly. *Neurology, 46,* 700–706.

Pfeiffer, E. (1975). A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *Journal of the American Geriatrics Society, 23,* 433–441.

Román, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., Amaducci, L., Orgogozo, J.-M., Brun, A., Hofman, A., Moody, D. M., O'Brian, M. D., Yamaguchi, T., Grafman, J., Drayer, B. P., Bennett, D. A., Fisher, M., Ogata, J., Kokmen, E., Bermejo, F., Wolf, P. A., Gorelick, P. B., Bick, K. L., Pajeau, A. K., Bell, M. A., DeCarli, C., Culebras, A., Korczyn, A. D., Bogousslavsky, J., Hartmann, A., & Scheinberg, P. (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology, 43,* 250–260.

Shay, K. A., Duke, L. W., Conboy, T., Harrell, L. E., Callaway, R., & Folks, D. G. (1991). The clinical validity of the Mattis Dementia Rating Scale in staging Alzheimer's dementia. *Journal of Geriatric Psychiatry and Neurology, 4,* 18–25.

Shepard, L., Camilli, G., & Averill, M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criteria. *Journal of Educational Statistics, 6,* 317–375.

Smith, G. E., Ivnik, R. J., Malec, J. F., Kokmen, E., Tangalos, E., & Petersen, R. C. (1994). Psychometric properties of the Mattis Dementia Rating Scale. *Assessment, 1,* 123–131.

Swaminathan, H., & Rogers, H. J. (1990). Detecting diferential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27,* 361–370.

van Belle, G., Uhlmann, R. F., Hughes, J. P., & Larson, E. B. (1990). Reliability of estimates of changes in mental status test performance in senile dementia of the Alzheimer type. *Journal of Clinical Epidemiology, 43,* 589–595.

Vangel, S. J., & Lichtenberg, P. A. (1995). Mattis Dementia Rating Scale: Clinical utility and relationship with demographic variables. *The Clinical Neuropsychologist, 9,* 209–213.

Vitaliano, P. P., Breen, A. R., Russo, J., Albert, M., Vitiello, M. V., & Prinz, P. N. (1984). The clinical utility of the Dementia Rating Scale for assessing Alzheimer's patients. *Journal of Chronic Diseases, 37,* 743–753.

Welsh, K. A., Fillenbaum, G., Wilkinson, W., Heyman, A., Mohs, R. C., Stern, Y., Harrell, L., Edland, S. D., & Beekly, D. (1995). Neuropsychological test performance in African-American and white patients with Alzheimer's disease. *Neurology, 45,* 2207–2211.

Woodard, J. L., Salthouse, T. A., Godsall, R. E., & Green, R. C. (1996). Confirmatory factor analysis of the Mattis Dementia Rating Scale in patients with Alzheimer's disease. *Psychological Assessment, 8,* 85–91.

Received June 6, 1997
Accepted January 20, 1998

Downloaded from https://academic.oup.com/psychsocgerontology/article/53B/6/P370/618616 by guest on 21 August 2022