

Education Bias in the Mini-Mental State Examination

RICHARD N. JONES AND JOSEPH J. GALLO

ABSTRACT. Education is correlated with cognitive status assessment. Concern for test bias has led to questions of equivalent construct validity across education groups. Following the work of previous researchers, we submitted Mini-Mental State Examination (MMSE) responses to external validation analyses. Subjects were older participants in the Epidemiologic Catchment Area study (age 50-98). Little evidence for test bias against those with low education was found. The correlation of MMSE scores and age was equivalent across high- and low-education groups ($-.29$ vs. $-.27$, $p = .48$), as was the correlation of MMSE scores and activities of daily living (ADL) functioning ($-.23$ vs. $-.27$, $p = .42$). The MMSE displayed significantly higher internal consistency reliability in the low-education group ($.75$ vs. $.72$, $p = .04$). The MMSE did not predict functional decline over 1 year or mortality over 13 years differently by level of educational attainment. Evidence for sex bias was found. The MMSE was more highly correlated with age among women than among men ($-.28$ vs. $-.21$, $p < .001$). The MMSE was more highly correlated with ADL impairment among women than among men ($-.30$ vs. $-.17$, $p = .01$). The MMSE predicted mortality differently according to participant sex ($p = .053$). The lack of evidence for bias provides little support to proposals to adjust MMSE scores according to level of education.

The correlation of education with performance on mental status tests is one of the more robust findings in studies of cognition and aging. Some authors have concluded that education may be the most significant risk factor for dementia identified to date (Mortimer & Graves, 1993). Others have argued that assessment may

be biased by education (Anthony et al., 1982; O'Connor et al., 1991). Twelve years ago, Jorm and colleagues (1988) published analyses addressing Mini-Mental State Examination (MMSE) bias by level of educational attainment among a sample of older adults using methods outlined in Berk's (1982) *Handbook of Methods for Detecting Test Bias*. Schmand and colleagues replicated these analyses in 1995 using data from the AMSTEL project.

The present analysis is a replication of the line of inquiry initiated by Jorm and colleagues (1988). The underlying conceptual model is consistent with the notion of construct validity described by Cronbach and Meehl (1955), and represents an attempt to move beyond the demonstration

From Hebrew Rehabilitation Center for Aged, Research and Training Institute, Boston, Massachusetts, US (R. N. Jones, ScD), and Department of Family Practice & Community Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, US (J. J. Gallo, MD, MPH).

Offprints. Requests for offprints should be directed to Richard N. Jones, ScD, HRCA Research and Training Institute, 1200 Centre Street, Boston, MA 02131, US. email: jones@mail.hrca.harvard.edu

of mean score differences but to show if and how the MMSE (Folstein et al., 1975) relates to external variables differently across groups defined by education. Our goal is to understand the degree to which the MMSE measures a characteristic of people presumed to be reflected in test scores. This characteristic—cognitive status—is expected to be related to certain external variables such as age, level of physical functioning, and mortality in a characteristic way. Further, if the test is valid in groups defined by high and low education, relationships with external variables should be equivalent across groups. For the test score to be considered free from bias in relation to specific external criteria, two individuals with the same MMSE score—one with high education, the other with low education—should have the same expected value on the external criteria (e.g., functional impairment) after controlling for the main effect of group membership. If the relationship of assessed cognition is functionally or substantially different by level of education (i.e., a significant group by MMSE score interaction), then we have evidence of a threat to the validity of the assessed score for that external criteria according to group membership (Jorm et al., 1988; Reynolds, 1982b).

We replicate the following tests of bias performed by Jorm and colleagues (1988): (a) equivalent prediction of activities of daily living (ADL) functional ability, (b) equivalent correlation with age, and (c) equivalent accuracy of measurement (internal consistency reliability). The current work expands upon previous research by adding (d) the prediction of functional decline and (e) the prediction of death. With the exception of internal consistency reliability, we do not consider item-level comparisons across groups here (e.g., factor loadings, rank order of item difficulty).

We address item-level issues elsewhere using modern psychometric techniques drawn from the educational and psychological testing literature (Jones, 1997; Jones & Gallo, 2000). Another way we extend previous work is in our treatment of education as a study variable. Education is considered in two different ways, one that takes into consideration age, sex, and ethnicity cohort differences in educational attainment when making group assignments. In addition, sex comparisons are included to provide a frame of reference for results obtained for education. We examined MMSE responses for bias by education and sex separately to allow the reader to gauge the importance of statistical evidence of test bias revealed for education groups by comparing to statistical results obtained for men and women. We hypothesized that the MMSE would relate to external variables equivalently and display equivalent internal characteristics across education and sex groups.

MATERIALS AND METHODS

Study Sample

The study sample was drawn from respondents to the five-site National Institute of Mental Health (NIMH) Epidemiologic Catchment Area (ECA) study conducted at Yale University, Johns Hopkins University, Duke University, UCLA, and Washington University. The methods and sample selection procedures used in the ECA study have been reported elsewhere (Eaton & Kessler, 1985; Regier et al., 1984). At each ECA site, investigators used survey sampling techniques to draw a representative sample of residents living within official Mental Health Catchment areas.

Sites taken together had a demographic pattern similar to that of the United States as a whole in 1980 (Leaf et al., 1991). This project was approved by the Committee on Human Research of the Johns Hopkins University School of Public Health.

Variables Under Study

Cognitive Assessment. At each ECA study site, the interview included the MMSE. For the purposes of the current study, responses that were not correct were considered errors. A summary MMSE score was obtained by summing the number of correct responses. Both the serial-seven subtraction item and the "world"-spelled-backwards item were included in the summary score, resulting in a maximum possible score of 35 (noted as MMSE35). This method of scoring the MMSE follows that of Jorm and colleagues (1988).

Low Education. We used two approaches to describe educational attainment. One, a mean split indicator, defined low education as fewer than the overall sample mean years of completed education (10 years). The second approach, a relativistic indicator, involved stratifying the sample by age at initial interview (grouped by years 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+), sex, and self-described ethnicity (White, African American, and other). Subjects with fewer than the median years of completed education within each of the resulting 48 strata were considered to have low education. The relativistic indicator attempts to account for the influence of social and cohort factors in making assignments to high- and low-attainment groups.

Ethnicity. Information on self-reported ethnicity was obtained by asking the following question: "Would you please look at this card and give me the letter of the

group that best describes your racial background?" The respondent then selected from a list: American Indian, Alaskan Native, Asian, Pacific Islander, Black-Not Hispanic, Hispanic, White-Not Hispanic. From responses to this question, we examined the distribution of years of completed education within age and sex strata for African Americans, Hispanics, and Whites and others to code the relativistic educational attainment indicator.

Functional Ability. Data regarding difficulty in a number of ADL were available for the Baltimore ECA site only. The summary ADL scale used in this analysis represents the number of eight areas (transferring from bed or chair, dressing, eating, bathing and toileting, carrying a bag of groceries across a room, bending, and cutting toenails) in which the subject indicated he or she had at least some difficulty. Choice of items was based on results of factor analyses of a broader array of ADL/instrumental activities of daily living (IADL) items. Functional decline was defined as any value at Wave 2 (1982) implying worse functioning than at Wave 1 (1981).

Death. Mortality data were collected at the third wave (13-year) follow-up of the Baltimore ECA cohort. Tracing procedures used to locate Wave 1 sample members have been described previously (Badawi et al., 1999).

Analytic Strategy

Correlation of MMSE With Age. Under the assumption that cognitive skills decline with age, the correlation of MMSE with age is an index of the test's validity. Under the hypothesis of equal validity across groups, the correlation of MMSE35 scores and age should be equivalent for men and for women and for high- and

low-education groups. To test this hypothesis, we estimated the correlation of MMSE35 scores and age for men and women and high- and low-education groups, which were compared using Fisher's z -transformation (Snedecor & Cochran, 1989).

Correlation of MMSE With Functional Ability. Similarly, functional limitations and cognitive performance are expected to be correlated. This correlation is also an index of the validity of the test. Unlike Jorm and colleagues (1988) and Schmand and colleagues (1995), we used a correlation approach rather than a regression approach to describe this relationship. The rationale for this analytic approach is addressed in our discussion. The strength of association between MMSE35 score and functional ability was tested by comparing Fisher's z -transformed correlation coefficients across groups.

Predicting Functional Decline. We examined the relationship of MMSE score to functional decline longitudinally. A series of logistic regression models were estimated to test the hypothesis that the MMSE35 score was not related differently to functional decline according to group membership (education, sex). The first two models fit a constant and the MMSE35 score, respectively, to the logit of the probability of functional decline over 1 year. The third model added a group indicator (one for either the mean split educational attainment indicator, the relativistic educational attainment indicator, or the sex indicator), to test the hypothesis of no group differences in the odds of functional decline, controlling for MMSE35 score at baseline. In the fourth model we added an interaction term between the group indicator and MMSE35 score to Model 3. If this parameter was significant (change in $-2\log$ likelihood, $p < .05$), the hypothesis of no difference in how

MMSE35 score relates to functional decline across groups was rejected.

Predicting Mortality. Logistic regression models were used to investigate possible group differences in the association of MMSE35 and mortality over 13-year follow-up in a manner similar to the prediction of functional decline. A significant interaction term indicated MMSE35 score related to mortality differently by group membership.

Internal Consistency Reliability. Part of the definition of construct validity is that the test measures the same domain with equal accuracy across groups. Bias exists in regard to construct validity when a test is shown to measure the same trait but with differing degrees of accuracy (Reynolds, 1982a). Cronbach's alpha (α) is a statistical summary of the internal consistency of a test. Groups were compared with a variance ratio statistic $(1 - \alpha_1)/(1 - \alpha_2)$, where α_1 is the larger of the internal consistency reliability estimates for the two groups, and α_2 the smaller of the two. The statistic follows an F distribution with N_1, N_2 degrees of freedom (Reynolds, 1982b). For the purposes of computing alpha, we used 27 MMSE items, treating each response as a dichotomous item with the exception of "world" spelled backwards and serial sevens, each of which was recorded with a single dichotomous indicator. This strategy avoids spuriously high ratings of reliability (Teresi & Holmes, 1994).

RESULTS

Sample Characteristics

The total number of respondents in the five-site ECA program numbered 20,861. Exclusions for the current analysis include those less than 50 years of age at first interview ($n = 10,594$) or missing age ($n =$

28), those not living at home ($n = 1,294$), or did not at least start the MMSE ($n = 326$). Respondents without a valid response to the question "What is the highest grade in school or year of college that you completed?" were excluded ($n = 63$). The final sample for analysis included 8,556 respondents. Only participants from the Baltimore ECA site had data available for analyses involving mortality ($n = 1,565$), ADL functioning ($n = 1,526$), and ADL decline ($n = 1,146$). Characteristics associated with nonresponse have been reported previously (Eaton et al., 1992). Subject characteristics are summarized in Table 1.

Correlation of MMSE With Age and ADL

The Pearson's correlation of age and MMSE35 score was not significantly different across education groups. The association of age and MMSE35 score was stronger for women ($r = -.28$) than for men ($r = -.21$, $p < .001$). These results are summarized in Table 2. The association of functional impairment and MMSE35 score was stronger among those with less than the overall sample mean level of educational attainment ($p = .053$), also summarized in Table 2. However, the relativistic educational attainment did not produce the same result ($p = .422$). Likely, this discrepancy is due to the association of raw years of education with age. The analysis regarding sex demonstrated that the strength of the association of MMSE35 score and functional ability was much stronger among women than among men ($p = .013$).

Predicting Functional Decline

Table 3 provides summary statistics for the logistic regression models predicting functional decline based on educa-

tional attainment and sex. Only summary fit statistics are provided. Our intent was limited to determine if the relationship of MMSE35 and functional decline was different across groups. Although MMSE35 score and educational attainment were both significantly associated with functional decline, the interaction term was not significant for both the mean split and relativistic indicators. Sex was not associated with functional decline over 1-year follow-up. We found no difference in the association of functional decline and cognitive impairment by sex.

Predicting Mortality

Both educational indicator models revealed no difference in the association of mortality across MMSE35 score by level of education (Table 4). This was not true for models with the sex covariate. A borderline significant MMSE35-by-sex interaction in predicting mortality over the 13-year interval was detected ($p = .053$). Within the range of MMSE35 scores where 95% of the community respondents scored (22 or higher), men faced a greater risk of death, but the risk of death increased more sharply with increasing cognitive impairment for women than for men.

Internal Consistency Reliability

The Cronbach's alpha internal consistency reliability estimate for the mean split educational attainment indicator group with high education was $\alpha_H = .66$, which was significantly lower than that for the group with low education ($\alpha_L = .75$; $F = 1.35$, $p < .001$). High- and low-education groups defined by the relativistic indicator also displayed significant differences in internal consistency according to education ($\alpha_H =$

TABLE 1. Respondent Characteristics, Sample for Analysis. Five-Site Collaborative Epidemiologic Catchment Area (ECA) Study and Baltimore ECA Site, 1981–1984 (Age 50–98 Years)

	Five-Site ECA Sample		Baltimore ECA Site	
	<i>n</i>	%	<i>n</i>	%
Respondent characteristic				
Total	8,556	100.0	1,531	100.0
ECA site				
Yale University, New Haven, CT	3,063	35.8	0	0.0
Johns Hopkins University, Baltimore, MD	1,531	17.9	1,531	100.0
Washington University, St. Louis, MO	1,082	12.6	0	0.0
Duke University, Raleigh-Durham, NC	2,031	23.7	0	0.0
UCLA, Los Angeles, CA	849	9.9	0	0.0
Age group at baseline interview, years				
50–54	870	10.2	195	12.7
55–59	1,010	11.8	222	14.5
60–64	1,297	15.2	258	16.9
65–69	1,928	22.5	312	20.4
70–74	1,470	17.2	258	16.9
75–79	1,035	12.1	166	10.8
80–84	605	7.1	75	4.9
85+	342	4.0	45	2.9
Self-described ethnicity				
American Indian	72	0.8	27	1.8
Alaskan Native	1	0.0	1	0.1
Asian	35	0.4	0	0.0
Pacific Islander	9	0.1	3	0.2
African American–Not Hispanic	1,582	18.5	347	22.7
Hispanic	342	4.0	11	0.7
White–Not Hispanic	6,450	75.4	1,142	74.6
Other	17	0.2	0	0.0
Not specified	48	0.6	0	0.0
Sex				
Women	5,273	61.6	962	62.8
Men	3,283	38.4	569	37.2
Level of education, highest grade completed				
None	112	1.3	16	1.0
1–5 (elementary school)	862	10.1	164	10.7
6–7 (some middle school)	1,001	11.7	262	17.1
8 (completed middle school)	1,388	16.2	301	19.7
9–11 (some high school)	1,653	19.3	350	22.9
12 (completed high school)	1,653	19.3	255	16.7
Grade 13 or higher (postsecondary)	1,887	22.1	183	12.0

.72, $\alpha_L = .75$, $F = 1.14$, $p = .04$). Internal consistency reliability was equivalent by sex ($\alpha_F = .75$, $\alpha_M = .75$, $F = 1.01$, $p = .44$).

DISCUSSION

Our analyses failed to demonstrate evidence for differential relationships of MMSE scores to external variables by level

of education, although the internal consistency reliability was higher among those with less education. These findings were true for education groups defined by the sample mean or by the median years of education within age, sex, and ethnicity strata. The same was not true for sex. The MMSE was more strongly related to age and functional disability among women. Thus, although there is considerable evi-

TABLE 2. Correlation of MMSE35 Score With Age and ADL Functional Impairment Within Education and Sex Groups. Epidemiologic Catchment Area Study (1981–1982) ($n = 8,556$ for Age Comparison, $n = 1,526$ for ADL Comparison)

External Characteristic	Educational Attainment Group								
	Relative to Overall Sample Mean ^a			Relative to Sex, Age, Ethnicity Cohort ^b			Sex		
	Low	High	p^c	Low	High	p^c	Women	Men	p^c
Age (50–98 years)	-.19	-.19	.796	-.27	-.29	.476	-.28	-.21	.001
ADL impairment (0–8)	-.26	-.16	.053	-.27	-.23	.422	-.30	-.17	.013

Note. Abbreviations defined in text. ^aLow education indicates fewer than 10 years of completed education. ^bLow education indicates fewer than the median years of education for their age, sex, and ethnicity strata. ^c p value obtained from comparison of the difference of Fisher's z-transformed correlation coefficients (Snedecor & Cochran, 1989).

TABLE 3. Summary of Model Fit, Logistic Regression Models of Functional Decline Over Approximately 1 Year. Baltimore Epidemiologic Catchment Area Program ($n = 1,146$)

Model Parameters	-2×Log-Likelihood	Improvement χ^2 (df)	p
0 Constant	777.09		
1 (Model 0) + MMSE35	764.57	12.51 (1)	<.001
Relativistic educational attainment (LOWEDUC) model			
2 (Model 1) + LOWEDUC	758.36	6.21 (1)	.013
3 (Model 2) + MMSE35 × LOWEDUC	758.07	0.29 (1)	.592
Mean split educational attainment (LESSMEAN) model			
4 (Model 1) + LESSMEAN	756.29	8.29 (1)	.004
5 (Model 4) + MMSE35 × LESSMEAN	755.76	0.52 (1)	.469
Sex model			
6 (Model 1) + MALE	763.66	0.92 (1)	.339
7 (Model 6) + MMSE35 × MALE	763.29	0.37 (1)	.543

TABLE 4. Summary of Model Fit, Logistic Regression Models of Death Over Approximately 13 Years. Baltimore Epidemiologic Catchment Area Program ($n = 1,519$)

Model Parameters	-2×Log-Likelihood	Improvement χ^2 (df)	<i>p</i>
0 Constant	2,096.45		
1 (Model 0) + MMSE35	2,048.279	48.17 (1)	<.001
Relativistic educational attainment (LOWEDUC) model			
2 (Model 1) + LOWEDUC	2,046.58	1.70 (1)	.192
3 (Model 2) + MMSE35 × LOWEDUC	2,046.07	0.51 (1)	.477
Mean split educational attainment (LESSMEAN) model			
4 (Model 1) + LESSMEAN	2,043.32	4.96 (1)	.026
5 (Model 4) + MMSE35 × LESSMEAN	2,043.19	0.13 (1)	.719
Sex model			
6 (Model 1) + MALE	2,024.32	23.95 (1)	.000
7 (Model 6) + MMSE35 × MALE	2,020.58	3.75 (1)	.053

dence to suggest the MMSE is not construct valid by sex, there is little support to the position that it is not construct valid by education.

Limitations of the current analysis should be discussed before we place our results in the context of previous research. One limitation relates to treatment of education as a study variable. Our sample was not large enough to group subjects into education quartiles or tertiles and maintain the conceptual advantages of considering relative educational attainment. Our sample includes older adults covering a broad age range, which might attenuate measures of association due to ceiling effects. In fact, the lower point estimate for the correlation of ADL and MMSE35 score among those with higher than the sample mean years of education is likely due to floor/ceiling effects on both measures. On the other hand, the broad age range limits attenuation due to restricted range and increases statistical power by providing a large sample size. Statistical power is of concern when conclusions are drawn from

null findings. The Baltimore sample was large enough to detect significant differences in correlation coefficients between about .10 and .14, and the five-site sample large enough to detect significant differences in correlation coefficients in the range of about .04 to .06 (assuming Type I error rate of 5%, Type II error rate of 20%, and equal-sized subgroups). We note that these ranges describe relatively small effects (Cohen, 1969), and that we are unaware of rules of thumb or accepted standards as to how large an effect should be to justify total score adjustment.

Other limitations include the choice of statistical model to relate variables that violate assumptions of the linear model. Unequal means and variances in MMSE score and external variables such as physical functioning confound the subtle cross-group differences we are attempting to reveal. A limitation specific to the mortality analysis is that cognitive impairment and low education have both been positively related to loss to follow-up and death in this cohort (Badawi et al., 1999). If many

TABLE 5. Comparison of Studies Conducting MMSE Item Bias Studies

Study	Sample	Correlation of MMSE With Age		Correlation of MMSE With ADL Functioning		Internal Consistency Reliability				
		Low Educ.	High Educ.	Low Educ.	High Educ.	Low Educ.	High Educ.			
Jorm et al. (1988)	Community-dwelling sample, age 70+; <i>n</i> = 123 at least some secondary education, <i>n</i> = 146 primary education only	-.24	-.21	.73	-.33	-.27	.47	.65	.56	.10
Schmand et al. (1995)	AMSTEL study: community-dwelling sample aged 65-84, high- and low- education groups matched on age, sex (<i>n</i> = 3,070)	-.20 (entire sample) ^c		.22	.36	<.001	.75 (entire sample) ^d			
Jones & Gallo (this study)	ECA study: community-dwelling sample aged 50-98; low education defined within age, sex, and ethnicity strata (<i>n</i> = 8,556 age and reliability, <i>n</i> = 1,531 ADL comparison)	-.27	-.29	.48	-.27	-.23	.42	.75	.72	.04

Note. Abbreviations defined in text.

^a*p* on differences in correlations transformed to Fisher's *z*; computed from published data where appropriate. ^bVariance ratio test. ^cGroup-specific correlations not reported, but significance of difference in Fisher's *z*-transformed values >.05. ^dGroup-specific alphas not reported, but reported to be not significantly different.

of those not located through the tracing procedures were in fact dead, our analyses would underestimate the mortality risk for those with low education and those with low cognitive functioning at baseline. No approach is perfect. Some tests confounded by floor/ceiling effects, others confounded by possible ascertainment bias, and many violate assumptions of the statistical model employed. The internal consistency reliability comparisons are confounded by base rate differences across groups (Teresi & Holmes, 1994), and the meaning of group differences not entirely clear (Loevinger, 1954). For this reason, like Jorm and colleagues (1988) and Schmand and coworkers (1995), we evaluated the validity of the MMSE across groups using several methods and several criteria, and looked for a convergence of various lines of evidence.

Our analyses represent a replication of work pursued using community samples in Hobart (Jorm et al., 1988) and Amsterdam (Schmand et al., 1995). Similar analyses were also reported by Murden and colleagues using a sample of primary care patients in New York City (Murden et al., 1991), although their focus was on ethnicity differences. We provide a comparison of results obtained by Jorm and colleagues (1988) and Schmand and coworkers (1995) and our own in Table 5. Considering three tests of MMSE construct validity by level of education (equivalent correlation with age, equivalent correlation with functional impairment, equivalent internal consistency reliability), in only two instances was there evidence of a threat to validity by level of education. In our research, we found that the internal consistency reliability of the MMSE was greater among those with low education. The meaning of this finding is unclear, and difficult to interpret in the light of base rate differenc-

es across groups (Loevinger, 1954; Teresi & Holmes, 1994). This finding was not reported by Jorm and colleagues (1988) and Schmand and colleagues (1995). Schmand and colleagues (1995) found significant differences in the slopes and intercepts of regression lines describing the dependency of physical functioning on MMSE score across education groups. This finding is not very compelling evidence of construct bias in the MMSE by level of education. The magnitude of differences in the model implied that functional impairment conditional on MMSE and group membership is very small. Furthermore, because functional impairment and MMSE summary scores tend to display skewed distributions in community surveys, ceiling effects and unequal variances threaten the validity of inferences drawn from regression (and correlational) approaches. High-education groups tend to score at the ceiling of the MMSE and thus have reduced variance on this measure. Differences in intercepts and slopes of regression lines relating functional impairment and MMSE scores across groups can be influenced by seemingly arbitrary decisions such as whether to model physical functioning as *impairment* (high scores implying more impairment) or *ability* (high scores implying more independence; see, for example, the different directions of correlation coefficients across studies in Table 5). For this reason we chose to model the association of ADL function and MMSE35 using a correlational rather than a regression approach.

One very important external criterion against which to assess the presence of education bias in mental status scores are neuropsychiatric diagnoses. Systematic misclassification of probable caseness by education, or selection bias, is an important problem for case identification activities.

The adequacy of attempts to reduce misclassification must be judged by the degree to which selection bias is reduced. If total score adjustment improves case identification, such a modification is justified. An excellent review of this topic as applied to cognitive screening in late life is provided by Kraemer and colleagues (1998), who argue cogently that published data indicate unadjusted scores are preferable when screening for dementia. Additionally, Lindeboom and colleagues (1996), in presenting further data from the AMSTEL project, conclude that adjustment is unlikely to improve case identification activities in representative samples. The rationale for adjusting total MMSE scores to remove the effect of education must therefore draw on evidence other than: that doing so improves case identification, or that the test has been shown to be biased.

REFERENCES

- Anthony, J. C., LeResche, L., Niaz, U., von Korff, M. R., & Folstein, M. F. (1982). Limits of the 'Mini-Mental State' as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*, 12, 397-408.
- Badawi, M. A., Eaton, W. W., Myllyluoma, J., Weimer, L. G., & Gallo, J. (1999). Psychopathology and attrition in the Baltimore ECA 15-year follow-up 1981-1996. *Social Psychiatry and Psychiatric Epidemiology*, 34(2), 91-98.
- Berk, R. A. (1982). *Handbook of methods for detecting test bias*. Baltimore: Johns Hopkins University Press.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Eaton, W. W., Anthony, J. C., Tepper, S., & Dryman, A. (1992). Psychopathology and attrition in the Epidemiologic Catchment Area surveys. *American Journal of Epidemiology*, 135, 1051-1059.
- Eaton, W. W., & Kessler, L. G. (1985). *Epidemiologic field methods in psychiatry: The NIMH Epidemiologic Catchment Area program*. New York: Academic Press.
- 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.
- Jones, R. N. (1997). *Construct validity and item analysis of the Mini-Mental State Examination by level of educational attainment and sex*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore, MD.
- Jones, R. N., & Gallo, J. J. (2000, May). Dimensions of the Mini-Mental State Examination among community dwelling older adults. *Psychological Medicine*, 30, 605-618.
- Jorm, A. F., Scott, R., Henderson, A. S., & Kay, D. W. (1988). Educational level differences on the Mini-Mental State: The role of test bias. *Psychological Medicine*, 18, 727-731.
- Kraemer, H. C., Moritz, D. J., & Yesavage, J. (1998). Adjusting Mini-Mental State Examination scores for age and educational level to screen for dementia: Correcting bias or reducing validity? *International Psychogeriatrics*, 10, 43-51.
- Leaf, P., Myers, J., & McEnvoy, L. (1991). Procedures used in the Epidemiologic Catchment Area study. In L. N. Robins & D. A. Regier (Eds.), *Psychiatric disorders in America* (pp. 11-32). New York: The Free Press.
- Lindeboom, J., Launer, L. J., Schmand, B. A., Hooyer, C., & Jonker, C. (1996). Effects of adjustment on the case-finding potential of cognitive tests. *Journal of Clinical Epidemiology*, 49, 691-695.
- Loevinger, J. (1954). The attenuation paradox in test theory. *Psychological Bulletin*, 51, 493-504.
- Mortimer, J., & Graves, A. (1993). Education and other socioeconomic determinants of dementia and Alzheimer's disease. *Neurology*, 43(Suppl. 4), S39-S44.

- Murden, R. A., McRae, T. D., Kaner, S., & Bucknam, M. E. (1991). Mini-Mental State exam scores vary with education in blacks and whites. *Journal of the American Geriatrics Society*, 39, 149-155.
- O'Connor, D. W., Pollitt, P. A., & Treasure, F. P. (1991). The influence of education and social class on the diagnosis of dementia in a community population. *Psychological Medicine*, 21, 219-224.
- Regier, D. A., Myers, J. K., Kramer, M., Robins, L. N., Blazer, D. G., et al. (1984). The NIMH Epidemiologic Catchment Area program. Historical context, major objectives, and study population characteristics. *Archives of General Psychiatry*, 41, 934-941.
- Reynolds, C. (1982a). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-208). New York: Wiley.
- Reynolds, C. R. (1982b). Methods for detecting construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 199-227). Baltimore: Johns Hopkins University Press.
- Schmand, B., Lindeboom, J., Hooijer, C., & Jonker, C. (1995). Relation between education and dementia: The role of test bias revisited. *Journal of Neurology, Neurosurgery, and Psychiatry*, 59, 170-174.
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods* (8th ed.). Ames, IA: Iowa State University Press.
- Teresi, J. A., & Holmes, D. (1994). Overview of methodological issues in gerontological and geriatric medicine. In M. P. Lawton & J. A. Teresi (Eds.), *Annual review of gerontology and geriatrics* (Vol. 14, pp. 1-22). New York: Springer Publishing.

Acknowledgments. Dr. Richard N. Jones's work on this project was supported in part by an NIMH Training Grant in Psychiatric Epidemiology at the Johns Hopkins University Department of Mental Hygiene, Professor William W. Eaton, Principal Investigator (5T32MH14592-22) and by NIH Grant AG17680, Richard N. Jones, Principal Investigator. Dr. Gallo is a Brookdale National Fellow in Geriatrics. Data gathering was supported by the Epidemiologic Catchment Area (ECA) program of the National Institute of Mental Health Division of Biometry and Epidemiology. The principal investigators and grant award numbers during data gathering were Jerome K. Myers at Yale University (MH34224), Morton Kramer at Johns Hopkins University (MH33870), and William Eaton at Johns Hopkins University for the Baltimore Follow-up (MH47447), Lee N. Robins at Washington University (MH33883), Dan Blazer and Linda George at Duke University (MH35386), and Richard Hough and Marvin Karno at the University of California, Los Angeles (MH35865). At the National Institute of Mental Health, principal collaborators during the data gathering were Darrel A. Regier, Ben Z. Locke, William W. Eaton, Carl A. Taupe, and Jack D. Burke, Jr.