Assessing Health Literacy in African American and Caucasian Adults: Disparities in Rapid Estimate of Adult Literacy in Medicine (REALM) Scores

Judy A. Shea, PhD; Benjamin B. Beers; Vanessa J. McDonald; D. Alex Quistberg; Karima L. Ravenell, MS; David A. Asch, MD, MBA

Background and Objectives: *The influence of literacy on health and health care is an important area of investigation. Studies with a literacy focus are most valuable when literacy is assessed with psychometrically sound instruments.* Methods: *This study used a prospective cohort sample of 1,610 primary care patients. Patients provided sociodemographics and took the Rapid Estimate of Adult Literacy in Medicine (REALM), a 66-item word pronunciation literacy test.* Results: *The sample was 65% African American; 66% were men; 51% had a high school education or less. REALM scores were significantly related to education, age, and race but not gender. When stratified by education, differences between African Americans and Caucasians remained significant. Using 19 different strategies to shorten the 66-item instrument, reliability coefficients above .80 were maintained.* Conclusions: *The REALM is a robust assessment of health literacy. However, the discordance in scores between African Americans and Caucasians with similar educational attainment needs to be further addressed. A much shorter instrument would still have internally consistent scores and potentially be more useful in clinical settings.*

(Fam Med 2004;36(8):575-81.)

Approximately 25% of adult Americans have limited literacy skills. For example, they are unable to complete a brief job application form or detect the time of a meeting from a brief schedule.[1] In recent years, low health literacy—the inability to read, understand, and use health care materials—has been shown to be related to poorer knowledge and understanding of one's health conditions.[2-5] Patients with lower literacy skills report lower rates of participation in preventive health services such as colorectal cancer screening.[6] Low literacy is also associated with worse health outcomes, such as glycemic control,[7] poorer health status,[8,9] and less satisfaction with health care,[10] even when controlling for other potentially confounding variables. The consistency of findings regarding literacy has helped make it the focus of several ongoing initiatives. The value of such studies will be highest when literacy is assessed with psychometrically sound instruments.

One of the most widely used instruments for studying literacy is the Rapid Estimate of Adult Literacy in Medicine (REALM), developed in the early 1990s by Davis and colleagues to help clinicians identify patients at greatest risk of having limited health literacy skills.[11] Early work with the REALM showed that scores compared favorably to other formal reading assessments and to assessments that test other skills (ie, comprehension), with correlation coefficients ranging from 0.80 to 0.90.[11-14] Early studies detailing development of the original 125-word version and the shortened 66-word version were each completed with slightly more than 200 patients, the majority of whom were African American. Little has been reported since then about how REALM scores vary by patient characteristics despite the fact that the REALM has been used in numerous studies. One exception is a recent abstract presenting a shortened (eight-item) version on the REALM from data based on 50 patients, suggesting that fewer items may be sufficient for screening.[15]

From the Center for Health Equity Research and Promotion (CHERP), Philadelphia Veterans Affairs Medical Center (Dr Shea, Ms McDonald, Ms Ravenell, and Dr Asch); the Department of Medicine (all) and the Leonard Davis Institute of Health Economics (Drs Shea and Asch), University of Pennsylvania.

Our research evaluated the validity and reliability of REALM scores among various patient subgroups. Specifically, for construct validity, we hypothesized that disparities in total REALM scores would be observed for subgroups of patients defined by education, gender, age, and race in a large and varied patient sample.[16] We further hypothesized that stratifying patients by education would eliminate any observed differences related to these sociodemographic characteristics, given that literacy is closely linked to educational attainment, generally falling three to five grade levels behind formal educational attainment.[17,18] We started with the expectation, drawn from testing theory, that within groups of similar education, there should not be consistent differences in performance related to test-taker demographics. If consistent differential item performance is observed, it might be a sign of item bias.[19]

For reliability we evaluated the internal consistency of multiple shortened forms of the REALM. We hypothesized that internal consistency would be maximized when assessed using all 66 items. However, given that earlier work with a shortened eight-item REALM reported coefficients above .90 with samples of just 50 patients,[15] we anticipated finding reasonable strategies to shorten the instrument without significantly compromising the internal consistency of the scores.

Methods

Subject Selection

Patients were recruited in primary care waiting areas at the Philadelphia Veterans Affairs Medical Center (VAMC) and three primary care clinics at the University of Pennsylvania Health System (UPHS). Patients were part of a larger study exploring literacy and patient satisfaction. The VAMC has a large clinic, seeing about 1,700 patients per month. They are 95% men, 45% African American, and approximately 45% are older than 55 years. Patients represent all levels of income and education, though lower education and socioeconomic status (SES) are prevalent. Two of the UPHS clinics served patients who were predominantly African American (90%) and come from the local West Philadelphia area, which, overall, has a low socioeconomic level. The third UPHS clinic had a mixture of patients drawn from the West Philadelphia community along with some university staff, faculty, and some patients traveling to this site from the suburbs. Overall, 40% of the clientele of the third UPHS clinic was African American.

A research assistant approached patients in the waiting area and invited them to participate in the study, explaining that it would take 15–20 minutes to complete all instruments and that responses would be anonymous. Patients were told that we were studying what patients liked and disliked most about making and having visits with their care providers. As part of the study, we asked them to complete several questionnaires, one of which was a reading exercise. Eligibility criteria included being at least 18 years old and able to speak English.

Over the course of the study, five research assistants collected data. One was a masters-trained geriatrician with extensive research experience, two were enrolled in the post-baccalaureate program prior to medical school, and two were in their final year of college in premedical majors. All were trained prior to data collection. They worked in pairs, and the pairs were regularly rotated. Rates for recruitment and completion were monitored. All patients we approached provided sociodemographic information (eg, age, ethnicity, race, educational attainment) via an oral interview.

Those who agreed to participate were given their choice of a tote bag or a $10 certificate to a local supermarket for their participation. Approximately 85% of the patients approached at each site agreed to participate in the study.

The study was approved by the Institutional Review Boards at both the University of Pennsylvania and the Philadelphia Veterans Affairs Medical Center. The study was explained to the potential participants following a script, and consent was obtained orally. Data were collected from May 2001 to April 2002.

Instruments

All participants in the study were tested with the REALM test, a word pronunciation test. The 66 medical words are ordered by difficulty, starting with one-syllable words and ending with multisyllable words. Subjects read as many words as they can. When they come to one they do not know, they are instructed to look at the rest of the words and pronounce any they can. Standard dictionary pronunciation is the scoring standard. The number of words read correctly is recorded, and this sum is translated to one of four grade-level literacy estimates. The 66-item REALM takes 2–3 minutes to administer and score.

Statistical Analyses

Data were analyzed with SAS v8.2 (Copyright (c) 1999–2001 by SAS Institute Inc, Cary, NC). Analysis of variance (ANOVA) and *t* tests were used to assess overall scale score and item performance (ie, the proportion of respondents answering the item correctly) for the demographically defined subgroups (eg, by age, education, gender, and race), and post hoc comparisons of means were performed with the Duncan test. Effect sizes were calculated to summarize differences in group performances. By convention, effect sizes of .20 are interpreted to be small, .50 are medium, and .80 are large.[20] Because the REALM is often presented as an ordinal score with four categories ranging from inadequate to adequate literacy, we also compared score

distributions between subgroups of patients using the chi-square statistic. Performance values (proportion correct) were the endpoints in most analyses. Samples of 400 in each comparison group would provide 80% power to detect an effect size of .10 in the middle of the score distribution at α=.05). Smaller samples would be required at the extremes.

To assess reliability or internal consistency, Cronbach's alpha was computed for various patient subgroups and subsets of items. Internal consistency is a coefficient that summarizes the extent to which items within an instrument assess a single domain. Coefficients can range from zero to one, with higher coefficients indicating greater homogeneity. The magnitude of the coefficient is a function of both the number of items in a scale and the average inter-item correlation. The goal is to select a set of items that produce a stable and high coefficient without introducing redundancy and/or creating respondent burden. Typically, coefficients of .80 are desired when studying group differences.[21]

To assess the reliability and degree of redundancy (and thus infer how the REALM might be shortened), we applied multiple item reduction strategies to two sets of respondents: a test sample and a validation sample, each of which was randomly selected. Similar results in two independent samples lend credibility to the results. We first examined coefficients for 13 samples of items: even-numbered items; odd-numbered items; random samples of 5, 10, 15, 20, and 25 items; and six systematic nonoverlapping samples of 11 items, choosing every seventh item beginning with the first, then the second, etc. Then we tried five reduction strategies that were based on rules developed for the test sample and applied to the validation sample: items with item-total correlations >.60, items with item-total correlations >.65, items with the effect size <.20 when comparing differences betweens African Americans and Caucasians, items with effect sizes <.10, and, finally, to maximize the consistency coefficient but minimize race differences, we selected items with item-total correlations >.60 and effect sizes <.20.

Results

A total of 1,805 patients participated in the study. We excluded the 192 (10%) who self-identified as other than African American or Caucasian (<1% American Indian, 1% Asian, 2% Hispanic, and 7% "other") and three who had incomplete REALM data. The remaining 1,610 patients are included in analyses. Demographic characteristics of the sample are presented in Table 1. REALM scores ranged from zero to 66. Two percent of the sample had estimated literacy skills below a third-grade level (scores of 0–18), while 59% were at the high school level or higher.

Table 2 reports mean REALM scores for multiple levels of education. As expected, REALM scores steadily and significantly increased with education level (P<.0001, effect size (ES)=1.24). Each of the four scores was significantly different from the others. Differences between REALM scores for males and females were not significant (P=.09, ES=.09). However, the mean REALM score for patients 65 and older was significantly lower than the mean scores for the two younger groups (P<.0001, ES=.25). The mean score of 55.7 for African Americans was significantly lower than the mean score of 61.0 for Caucasian patients (P<.0001, ES=.49).

Table 3 shows the results when stratifying by education and examining age and race. The differences in observed REALM scores among age groups were no longer significant once stratified by education level, except at the college level. However, differences between African Americans and Caucasians remained significant at all four education levels, with effect sizes between .32 and .48. African Americans scored lower on the REALM than Caucasians even when controlling for educational attainment.

The correlation of item difficulties between African Americans and Caucasians was .92, suggesting that the same items were difficult for both groups. However, in

Table 1

Characteristics of the Study Subjects

Group	*n**	*%*
Gender		
Male	1,052	66
Female	554	34
Age		
< 45 years	427	27
45–64 years	788	49
65+ years	395	25
Race		
African American	1,042	65
Caucasian	568	35
Education		
< High school	227	14
High school/GED	590	37
Some post high school	489	30
4-year college degree +	301	19
REALM		
0–18 (≤ third grade)	39	2
19–44 (fourth–sixth grade)	126	8
45–60 (seventh-eighth grade)	503	31
61–66 (≥ ninth grade)	942	59

* All categories do not sum to 1,610 because of missing data.

GED—General Educational Development
REALM—Rapid Estimate of Adult Literacy in Medicine

Table 2

Mean REALM Performance for Subgroups of Primary Care Patients

Group	*n*	*Mean (SD)**	P *Value*	*Effect Size***
Education				
< High school	227	48.3 (16.6)	< .0001	1.24
High school/GED	590	56.5 (11.1)		
Some college or technical school	489	60.0 (7.0)		
4-year college degree +	300	62.9 (6.0)		
Gender				
Male	1,052	57.2 (11.2)	.09	.09
Female	554	58.2 (11.2)		
Age				
<45 years	427	58.7 (10.4)*	< .0005	.25
45–64 years	788	57.9 (10.6)*		
65+ years	395	55.8 (12.9)		
Race				
African American	1,042	55.7 (11.8)	< .0001	.49
Caucasian	568	61.0 (9.0)		

* All group means within a sociodemographic variable are different from one another according to Duncan's post hoc comparison of means except for the two youngest age groups.

**Effect sizes are interpreted as .20 is small, .50 is medium, and .80 is large.

REALM—Rapid Estimate of Adult Literacy in Medicine
SD—standard deviation
GED—General Educational Development

Table 3

REALM Performance by Age and Race, Stratifying by Education

	< High School		*High School*		*Some College or Technical School*		*College +*	
	n	*Mean (SD)*	*n*	*Mean (SD)*	*n*	*Mean (SD)*	*n*	*Mean (SD)*
Age								
<45	33	47.6 (17.0)	163	56.4 (11.3)	129	60.3 (6.9)	101	64.2 (2.5)
45–64	90	49.5 (17.2)	301	56.9 (10.3)	268	59.8 (7.2)	127	62.2 (7.1)
65+	104	47.5 (15.9)	126	55.7 (12.8)	92	60.1 (6.9)	72	62.1 (6.9)
P value	.68		.58		.79		.02	
Effect size*	.01		.06		.03		.43	
Race								
African American	161	46.5 (16.7)	420	55.0 (11.6)	340	59.0 (7.5)	118	61.7 (5.3)
Caucasian	66	52.8 (15.5)	170	60.2 (8.7)	149	62.2 (5.3)	182	63.6 (6.2)
P value	.0098		<.0001		<.0001		.0054	
Effect size*	.38		.48		.46		.32	

* Effect sizes are interpreted as .20 is small, .50 is medium, and .80 is large.

REALM—Rapid Estimate of Adult Literacy in Medicine
SD—standard deviation

each of the 58 (of 66) items for which there was a significant difference in performance, African Americans performed more poorly than Caucasians. Table 4 reports the 10 items with the largest race disparity in performance. Looking at the position of the item, it can be seen that they are not exclusively at the difficult end of the scale. The middle two columns show the proportion in each group who pronounced the word correctly. The magnitude of these differences reflects moderately strong effect sizes. Looking at differences between African Americans and Caucasians on each item within each education level showed that the differences were pervasive rather than occurring among just a few items, suggesting they might be candidates for removal. Over all 66 items, performance differences were observed for 30% of the items in the group with less than a high school education, 68% of the items for the high school/GED group, 36% of the items in the group with some post high school, and 20% of the items in the group with at least a 4-year college degree.

The results for internal consistency showed more consistency across groups. For the total group of patients and the total 66-item REALM, Cronbach's alpha was .96, suggesting that the REALM provides very precise scores but with redundancy. The alpha was .92 or higher for each of the subgroups defined by all strata of education crossed by age, gender, and race. Results for the multiple-item reduction strategies are summarized in Table 5 for the randomly selected test and validation samples. With only 33 items in the odd or even halves, alphas remain above .90. Random samples of 5, 10, 15, 20, and 25 items gave increasingly higher coefficients. For six nonoverlapping samples of 11

Table 4

Proportion of Respondents Correctly Answering Items With Largest Race Differences*

Item	*Position***	*African American (n=1,042) Proportion Correct*	*Caucasian (n=568) Proportion Correct*	*Effect Size****
Colitis	48	47	82	.75
Osteoporosis	65	53	78	.53
Testicle	47	72	91	.48
Jaundice	25	67	87	.45
Seizure	18	72	90	.44
Bowel	19	72	90	.44
Pelvic	24	75	91	.43
Directed	44	71	92	.39
Fatigue	23	83	95	.39
Menstrual	46	78	92	.38

* All differences are statistically significant.
** Position refers to position within the list of 66 words, which move from easy to difficult.
*** Effect sizes are interpreted as .20 is small, .50 is medium, and .80 is large.

Table 5

Results of Multiple Strategies to Reduce the Length of the REALM

Strategy	*# of Items*	*Test Sample a*	*Validation Sample a*
All items	66	.96	.96
Odd items	33	.93	.93
Even items	33	.92	.92
Random	5	.56	.56
	10	.81	.80
	15	.84	.83
	20	.89	.89
	25	.91	.91
Systematic samples: every seventh item beginning with the:			
First item	11	.80	.80
Second item	11	.82	.81
Third item	11	.81	.79
Fourth item	11	.77	.78
Fifth item	11	.82	.81
Sixth item	11	.79	.78
Item-total correlation ≥ .60	21	.94	.93
Item-total correlation ≥ .65	11	.92	.90
Race effect size ≤ .20	26	.92	.91
Race effect size ≤ .10	7	.80	.74
Item-total correlation ≥ .60 and race effect size ≤ .20	4	.75	.69

items, the coefficients in both samples of patients were always near .80. Choosing the 21 items with the highest item-total correlation (>.60 in sample 1) gave coefficients above .90. The 11 items with item-total correlations greater than or equal to .65 had coefficients nearly as high. Second, taking into account the pervasive race differences and selecting the items with effect sizes less than or equal to .20 (26 items) or less than or equal to .10 (seven items) provided reasonably high coefficients. Finally, maximizing the consistency coefficient but minimizing race differences resulted in coefficients of .75 and .69 in the two samples using just four items.

Conclusions

This study is notable because of its relatively large size and its ability to investigate what may be separable associations among race, age, education, and health literacy. Overall, much of what we found was anticipated. As many others have observed, literacy is strongly related to education, though there is often a sizable gap between education level and literacy skill.[1,17,18,22] Literacy is also related to age and race in predictable ways, with younger patients and nonminority patients having better skills. However, the large and significant differences among age groups disappeared when we stratified by education.

As others have suggested, we also found that the REALM has substantial item redundancy.[15] While redundancy is good if one has time and needs very precise scores about individuals, however, little measurement information was lost by cutting down the number of items on the test.

Some of our observations in this study, however, were not anticipated and deserve further comment. Average REALM scores for African Americans were significantly lower than for Caucasians, even when compared within education strata, except for the college-educated groups. It is particularly notable that multiple items showed differential performance, and in every case the differences favored Caucasians. Others have found differences in literacy related to race and ethnicity, sometimes while controlling for education, but the task of sorting out how race is related to literacy assessment has had less breadth and depth, given other objectives of past work.

Why then were there so many differences related to race even when stratifying by education? First, the differences we found might be spurious, in that they might not be found again in repeat studies. Yet, this possibility seems unlikely considering the substantial statistical significance we observed. Second, the differences might reflect measurement error, in that not all educational experiences are equal. Thus, for example, completion of high school in one neighborhood school may not indicate the same level of educational attainment as completion of high school in a different school.

Although we did not directly assess educational quality, school systems vary in their resources and their success, and these differences might fall along racial lines. Third, differences might reflect confounding by unmeasured characteristics of the subjects also associated with their performance on the REALM. For example, perhaps unmeasured cultural or language differences are simultaneously associated with African Americans and lower REALM performance. Socioeconomic status is also a likely important covariate.

Limitations

This study has several limitations. First is the reliance on self-reported education, defined simply as the highest level of education completed. More information on the educational experience might help explain the differential item performance. The wider standard deviations in REALM performance for African Americans may be a sign of the great variability in their educational experiences. Second, the sample comprised volunteers in waiting areas of primary care clinics. Possibly those with lower literacy skills were more likely to refuse participation in our study or less likely to even be in a clinic. Third, the sample was restricted to those who self-identified as African American or Caucasian. Thus it is unclear how scores would perform for other groups of patients defined by ethnicity (eg, Hispanic, Asian) and/or those who do not have English as a first language. Fourth, the levels of health literacy observed in this study are relatively high compared to others, suggesting that spectrum bias may be influencing the results. Fifth, we did not measure visual acuity, thus we can not say how vision problems influenced literacy assessment.

Practical tools are needed if literacy is to become a routine part of health assessment.[23,24] By many standards the REALM performs extremely well. Scores are related as expected to education, age, and race. Although the instrument takes only 2 to 3 minutes to administer, our results suggest that the instrument can be considerably shortened, which may increase its practical use in both clinical and research settings. The original REALM authors, using input from subsequent users, could put forth a much shorter instrument. It would probably benefit the field to have a single shorter REALM rather than each investigator developing his or her own short list.

Another way to judge a tool is to ask if scores are comparable across groups who a priori might be expected to perform similarly, for example, groups with equivalent educational backgrounds. This was not the case, however, with the REALM. The persistent and unexplained differences in REALM scores associated with race might suggest bias in the test's content. The next step is to conduct studies that delve much more deeply into quality indicators of educational backgrounds. Should the findings persist, a strong argument could be made to revise the REALM, using only those items that do not show bias. The good news is that such items are present, and scores remain reliable even with a much smaller number of items. Naturally, in time, questions of bias would need to be extended to other races and ethnic groups, but doing so poses questions of the viability of a word pronunciation test with groups of patients who do not have English as a first language.

In addition to the language issues just mentioned, for health literacy researchers, these results point to several avenues for further study, such as item bias, instrument efficiency, instrument design for screening versus diagnostic tasks, and definition of global versus specific health literacy. For clinicians, these results provide evidence that many patients, but especially African American patients, are not familiar with words that may seem commonplace to clinicians. Making the time to check a patient's understanding should enhance the processes, and perhaps the outcomes, of care.

Acknowledgments: This research was supported by grant number PCC-98-071-1 from the Veterans Affairs Health Services Research and Development Office.

An earlier version of this work was presented at the May 2003 Annual Meeting of the Society of General Internal Medicine in Vancouver, BC, and the March 2003 Northeast Regional Meeting of the Society of General Internal Medicine in Baltimore.

Corresponding Author: Address correspondence to Dr Shea, University of Pennsylvania, 1223 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021. 215-573-5111. Fax: 215 573-8778. sheaja@mail.med.upenn.edu.

References

1. Kirsch I, Jungeblut A, Jenkins L, et al. Adult literacy in America; a first look at the results of the National Adult Literacy Survey. Washington, DC: Office of Educational Research and Development, US Department of Education, 1993.
2. Pearlman DN, Clark MA, Rakowski W, et al. Screening for breast and cervical cancers: the importance of knowledge and perceived cancer survivability. Women's Health 1999;28(4):93-112.
3. Lindau ST, Tomori C, Lyons T, et al. The association of health literacy with cervical cancer prevention knowledge and health behaviors in a multiethnic cohort of women. Am J Obstet Gynecol 2002;186(5):938-43.
4. Williams MV, Baker DW, Honig EG, et al. Inadequate literacy is a barrier to asthma knowledge and self-care. Chest 1998;114:1008-15.
5. Williams MV, Baker DW, Parker RM, et al. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med 1998;158(2):166-72.
6. Guerra CE, Dominguez F, Shea JA. Literacy, knowledge, and behavior toward colorectal cancer screening in Philadelphia's Hispanic community. J Gen Intern Med 2002;17(suppl):S160.
7. Schillinger D, Grumbach K, Piette J, et al. Association of health literacy with diabetes outcomes. JAMA 2002;288(4):475-82.
8. Weiss BD, Hart G, McGee DL, et al. Health status of illiterate adults: relation between literacy and health status among persons with low literacy. J Am Board Fam Pract 1992;5(3):257-64.
9. Guerra C, Shea JA. Functional health literacy, comorbidity, and health status. J Gen Intern Med 2003;18(suppl):S174.
10. Shea JA, Guerra C, Weiner J, et al. Health literacy and patient satisfaction. J Gen Intern Med 2003;18(suppl):S187-S188.
11. Davis TC, Long SW, Jackson RH, et al. Rapid Estimate of Adult Literacy in Medicine: a shortened screening instrument. Fam Med 1993;25(6):391-5.

12. Davis TC, Crouch MA, Long SW, et al. Rapid assessment of literacy levels of adult primary care patients. Fam Med 1991;23(6):433-5.
13. Parker RM, Baker DW, Williams MV, et al. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995;10:537-45.
14. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999;38:33-42.
15. Bass PF, Wilson JF, Griffith CH. A shortened instrument for literacy screening. J Gen Intern Med 2001;16:117.
16. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. Oxford, UK: Oxford University Press, 1995.
17. Hussey L, Gilliland K. Compliance, low literacy, and locus of control. Nurs Clin North Am 1989;24:605-11.
18. Wilson FL. Measuring patients' ability to read and comprehend: a first step in patient education. Nursing Connections 1995;8(4):17-25.
19. Hambleton RK, Swaminathan H. Item response theory: principles and applications. Boston: Kluwer-Nijhoff, 1985.
20. Murphy KR, Myors B. Statistical power analysis. A simple and general model for traditional and modern hypothesis tests. Mahwah, NJ: Lawrence Erlbaum Associates, 1998.
21. Nunnally JC Jr. Introduction to psychological measurement. New York: McGraw-Hill, 1970.
22. Reder S. The state of literacy in America: estimates at the local, state, and national levels. Washington, DC: National Institute for Literacy, 1998.
23. Davis TC, Michielutte R, Askov EN, et al. Practical assessment of adult literacy in health care. Health Educ Behav 1998;25:613-24.
24. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA 1999;281:552-7.