LETTERS

MEASURING ENGLISH PROFICIENCY AND LANGUAGE PREFERENCE: ARE SELF-REPORTS VALID?

A recent *Journal* article[1] found language preference was often unassociated, and English proficiency was consistently associated, with self-rated health—implying these should be modeled separately. A reexamination of this study suggests the reported conclusions depended on erroneous measures and highlights the importance of validity in linguistic studies.

Two alternative linguistic indicators, which have stronger face validity, were examined alongside the self-reports used in the original study in a replication using the same National Latino and Asian American Study (NLAAS) data. Interviewers rated respondents' English proficiency as non-English speaking, poor, fair, good, or excellent, and this was used to assess the criterion validity of self-reported language proficiency. Stronger preferences for English should be associated with conducting the survey in English; hence, the survey language measure was used to assess the predictive validity of self-reported language preference. A combined index was computed by standardizing interviewers' reports of English proficiency and conducting the interview in English—so each item weighed equally in the final index—and then summing the resulting quotients (Cronbach $\alpha=0.78$). These items were then tested in separate models predicting the 5 categorizations of self-rated health, replicating the original model specification, using ordinal logistic regression. (A complete methodological appendix is available as a supplement to the online version of this article at http://www.ajph.org.)

All interviewer-rated non-English speakers were misclassified; 75% reported they spoke poor, 18% fair, and 7% good English. This misclassification persisted where overlapping categorization between interviewer and self-reports existed. Among respondents with poor English proficiency, for example, 28% reported fair and 7% good English speaking proficiency. Overestimation tendencies also interfered with self-reported preferences. About 18% of non-English speakers reported using some English, 7% equal parts English and their native language, and 3% mostly English with friends. Respondents who reported speaking only English with friends versus mostly their native language had similar tendencies to interview in English (18% and 17%, respectively). The strong reliability originally reported with other self-reported measures for which no alternative existed, e.g., reading proficiency, suggests these may be similarly biased.

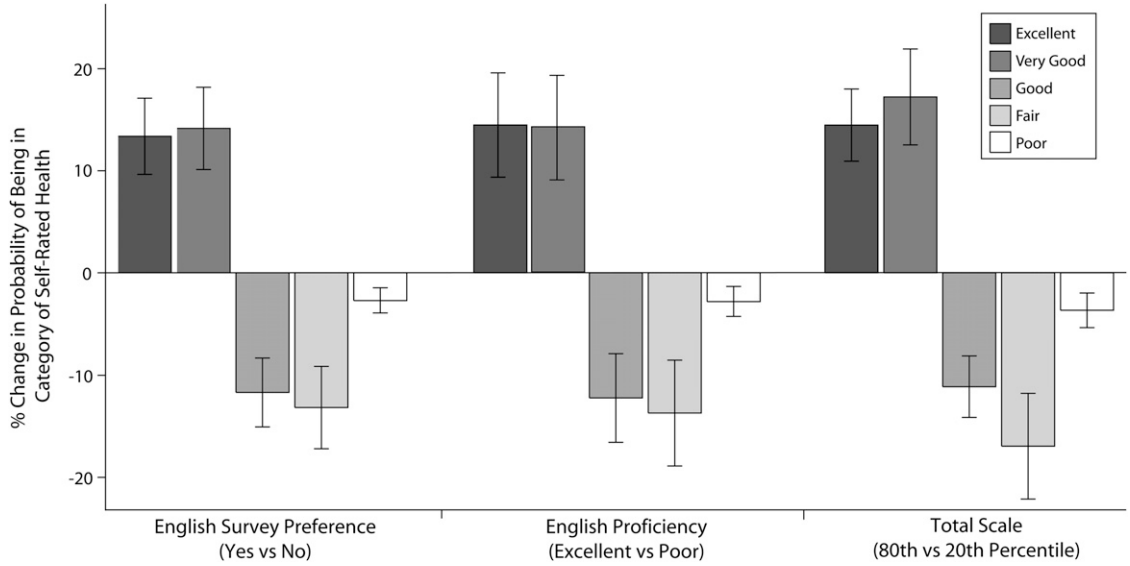
Behavioral measures of English proficiency, preference, and their combined scale, contrary to the original study, had similar significant associations with self-rated health) Figure 1). For example, surveying in English versus surveying in another language was associated with a 13% (95% confidence interval [CI]=10, 17) higher probability of excellent health and a corresponding 4% (95% CI=2, 5) lower probability of poor health. Standardized coefficients were computed to compare the relative effect size across measures, given the variation in measurement units. English proficiency, ($B=-0.57$; 95% CI=−0.73, −0.42), preference, ($B=-0.52$; 95% CI=−0.71, −0.34), and their combined scale ($B=-0.70$; 95% CI=−0.88, −0.52), had statistically indistinguishable associations with self-rated health.

It is well known that self-reports are susceptible to large systematic biases,[2] yet many studies have relied on self-reported linguistic measures. Among the 15 observational studies in the *Journal* since 1999 that used acculturation, acculturative, or acculturated in their title or abstract 10 (67%)[3–12] relied entirely on self-reported linguistic measures and only 3 (20%)[13–15] included observed linguistic traits. Public health has

Valérie Boyer has proposed legislation in France that would require all digitally altered photographs of people used in advertising to be labeled as retouched. Boyer, with a background in health administration and 2 adolescent daughters, became interested in the pressures on adolescents and young women to match the fashionable ideal of a thin body and perfect skin upon reflection as a mother. Photograph by Francois Guillot/Agence France-Presse. Printed with permission of Getty Images.

Letters to the editor referring to a recent Journal article are encouraged up to 3 months after the article's appearance. By submitting a letter to the editor, the author gives permission for its publication in the Journal. Letters should not duplicate material being published or submitted elsewhere. The editors reserve the right to edit and abridge letters and to publish responses.

Text is limited to 400 words and 10 references. Submit online at www.ajph.org for immediate Web posting, or at submit.ajph.org for later print publication. Online responses are automatically considered for print publication. Queries should be addressed to the Editor-in-Chief, Mary E. Northridge, PhD, MPH, at men11@columbia.edu.



Note. Each set of probabilities indicates a separate regression with adjustment for education, ethnicity, gender, region of the US in which respondents resided, age, and social desirability.

FIGURE 1—Change in probability of self-rated health by English language differences in survey preference, proficiency, and their combined scale.

been strongly critical of biomedical models[16]; however, when it comes to measures, a biomedical perspective in which the validity of survey measures are assumed appears common. The possible biases that interfered with the replicated study[1] may similarly impact other studies, especially those that relied entirely on self-reported measures. The reasons are many, but in addition to being a health determinant, English may be perceived as relevant to social status and respondents will be tempted to overestimate their English traits. Self-reports may also be limited because of poor across-subject reliability, where self-reports are under- or overestimated relative to direct observation.

It is admirable that Gee et al.[1] have conducted research on measurement issues in acculturative studies, but their conclusions are questionable given the systematic error in their primary measures. Future research should instead focus on the validity of linguistic measures by refining-observational measures that overcome the limitations of self-reports. Studies focused on enhancing the reliability and validity of interviewer observed linguistic measures are needed. Given the deviations of self-reports from face valid alternatives, it is advisable to apply similar strategies, as explored here, to other studies. ■

John W. Ayers, MA

About the Author

John W. Ayers is with the Department of Health, Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, and the Center for Behavioral Epidemiology and Community Health, Graduate School of Public Health, San Diego State University, San Diego, CA.

Correspondence should be sent to John W. Ayers; Department of Health Behavior and Society, Johns Hopkins Bloomberg School of Public Health, Hampton House, Baltimore MD, 21205; (e-mail: jayers@jhsph.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints/Eprints" link.

This letter was accepted February 15, 2010.

doi:10.2105/AJPH.2010.194167.

Acknowledgments

This work was supported by federal student work study aid and the National Institutes of Alcohol Abuse and Alcoholism (RC1-AA018970-01).

I would like to thank C. Richard Hofstetter, David Holtgrave, Melbourne Hovell, Carl Latkin, and Keith Schnakenberg for helpful comments.

Human Participant Protection

The Johns Hopkins Bloomberg School of Public Health's institutional review board approved this study.

References

1. Gee GC, Walsemann KM, Takeuchi DT. English proficiency and language preference: testing the equivalence of two measures. *Am J Public Health.* 2010;100(3):563–569.

2. Zaller J. *The Nature and Origins of Mass Opinion.* Cambridge, MA: Cambridge University Press; 1992.

3. O'Malley AS, Kerner J, Johnson AE, Mandelblatt J. Acculturation and breast cancer screening among Hispanic women in New York City. *Am J Public Health.* 1999;89(2):219–227.

4. Alderete E, Vega WA, Kolody B, Aguilar-Gaxiola S. Lifetime prevalence of and risk factors for psychiatric disorders among Mexican migrant farmworkers in California. *Am J Public Health.* 2000;90(4):608–614.

5. Crespo CJ, Smit E, Carter-Pokras O, Andersen R. Acculturation and leisure-time physical inactivity in Mexican American adults: results from NHANES III, 1988-1994. *Am J Public Health.* 2001;91(8):1254–1257.

6. Pérez-Stable EJ, Ramirez A, Villareal R, et al. Cigarette smoking behavior among US Latino men and women from different countries of origin. *Am J Public Health.* 2001;91(9):1424–1430.

7. Honda K. Factors associated with colorectal cancer screening among the US urban Japanese population. *Am J Public Health.* 2004;94(5):815–822.

8. Gomez SL, Kelsey JL, Glaser SL, Lee MM, Sidney S. Immigration and acculturation in relation to health and health-related risk factors among specific Asian subgroups in a health maintenance organization. *Am J Public Health.* 2004;94(11):1977–1984.

9. Wilkinson AV, Spitz MR, Strom SS, et al. Effects of nativity, age at migration, and acculturation on smoking among adult Houston residents of Mexican descent. *Am J Public Health.* 2005;95(6):1043–1049.

10. Detjen MG, Nieto FJ, Trentham-Dietz A, Fleming M, Chasan-Taber L. Acculturation and cigarette smoking among pregnant Hispanic women residing in

the United States. *Am J Public Health.* 2007;97(11):2040–2047.

11. Cruz GD, Chen Y, Salazar CR, Le Geros RZ. The association of immigration and acculturation attributes with oral health among immigrants in New York City. *Am J Public Health.* 2009;99(Suppl 1):S474–S480.

12. Shelley D, Fahs M, Scheinmann R, Swain S, Qu J, Burton D. Acculturation and tobacco use among Chinese Americans. *Am J Public Health.* 2004;94(2):300–307.

13. Brett KM, Higgins JA. Hysterectomy prevalence by Hispanic ethnicity: evidence from a national survey. *Am J Public Health.* 2003;93(2):307–312.

14. Vega WA, Sribney WM, Achara-Abrahams I. Co-occurring alcohol, drug, and other psychiatric disorders among Mexican-origin people in the United States. *Am J Public Health.* 2003;93(7):1057–1064.

15. Martin MA, Shalowitz MU, Mijanovich T, Clark-Kauffman E, Perez E, Berry CA. The effects of acculturation on asthma burden in a community sample of Mexican American schoolchildren. *Am J Public Health.* 2007;97(7):1290–1296.

16. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav.* 1995;35(extra issue):80–94.

GEE ET AL. RESPOND

Many studies use self-reported language measures.[1–3] We cautioned against conflating English proficiency and language preference.[4] We thank Ayers for confirming that English proficiency is related to self-rated health, regardless of whether proficiency is measured via self-report or interviewer assessment.

Ayers' main point is that objective indicators of proficiency are more accurate than self-report indicators,[2] a point we made in our paper.[4(p568)] Ayers argues that interviewer assessments should have been used as the measure of English proficiency because it is "objective" and face valid. As members of the team who collected the data, however, we caution against overstating these assessments. National Latino and Asian American Study (NLAAS) interviewers were well-trained, but they were not trained to assess English proficiency for research purposes. Hence, interviewer assessments have unknown biases.

Ayers' analysis of language preference differs from ours. We find that preference is inconsistently associated with self-rated health. Ayers claims this inconsistency is spurious, resulting from our inclusion of survey language. After removing survey language, he finds that preference measures are associated with self-rated health. He assumes that survey language measures preference.

We made a different assumption. We included survey language to control for instrumentation bias related to translation. Although NLAAS uses state-of-the-art translation methods, translations may still be nonequivalent. There is no "correct" assumption here, although if anything, our assumption made our analyses too conservative.

Nevertheless, our conclusions were derived after modeling binary and ordinal measures of self-rated health; Ayers considers only the latter. Table 1 replicates our study by modeling binary and ordinal measures of self-rated health, omitting survey language. Our ordinal analyses are consistent with Ayers', yet the binary analysis show that preference is not consistently related to self-rated health. Accordingly, a complete replication supports our original findings.

Moreover, one must not just analyze data, but also interpret it. The "unilinear" acculturation perspective predicts a gradient, with bilingual respondents falling intermediate between English and Asian speakers. In Ayers' Figure 1, the finding for language preference with friends is consistent with the unilinear perspective, but language of thinking or with friends is not. English speakers are not significantly different from non-English speakers. The only difference is between bilingual and non-English speakers. Ayers does not provide a substantive interpretation for these inconsistencies, but very similar results are discussed in our original article.[4(p567)]

We hope that researchers evaluate the central question of whether proficiency is equivalent to preference. Adding alternative

TABLE 1—Reanalysis of English Proficiency and Language Preference Measures: National Latino and Asian American Study, 2002–2003

	Binary Self-Rated Health, b (SE)	Ordinal Self-Rated Health, b (SE)
Language preference (continuous)		
When thinking	−0.139 (0.083)	−0.155** (0.056)
With family	−0.002 (0.091)	−0.035 (0.076)
With friends	−0.245* (0.093)	−0.190** (0.069)
Language preference when thinking (categorical)		
Asian language (Ref)	1.00	1.00
Both languages	−0.597** (0.199)	−0.582*** (0.154)
English	−0.302 (0.226)	−0.362* (0.176)
Language preference when speaking with family (categorical)		
Asian language (Ref)	1.00	1.00
Both languages	−0.269 (0.252)	−0.543* (0.209)
English	−0.002 (0.337)	0.018 (0.239)
Language preference when speaking with friends (categorical)		
Asian language (Ref)	1.00	1.00
Both languages	−0.794*** (0.204)	−0.535** (0.164)
English	−0.568* (0.233)	−0.496** (0.177)
Language preference scale	−0.175 (0.110)	−0.174* (0.079)
Combined scale	−0.592*** (0.132)	−0.493*** (0.082)

Note. Analyses are similar to Table 3 of Gee et al.,[4] except that survey language is not controlled. The sample size was n = 1639.