



# Measurement confounding affects the extent to which verbal IQ explains social gradients in mortality

## Citation

Chapman, Benjamin, Kevin Fiscella, Paul Duberstein, Ichiro Kawachi, and Peter Muennig. 2014. "Measurement Confounding Affects the Extent to Which Verbal IQ Explains Social Gradients in Mortality." *Journal of Epidemiology and Community Health* 68 (8): 728–33. <https://doi.org/10.1136/jech-2013-203741>.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41275567>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)



# HHS Public Access

Author manuscript

*J Epidemiol Community Health*. Author manuscript; available in PMC 2016 April 26.

Published in final edited form as:

*J Epidemiol Community Health*. 2014 August ; 68(8): 728–733. doi:10.1136/jech-2013-203741.

## Measurement confounding affects the extent to which verbal IQ explains social gradients in mortality

Benjamin Chapman<sup>1</sup>, Kevin Fiscella<sup>2,3</sup>, Paul Duberstein<sup>1,2</sup>, Ichiro Kawachi<sup>4</sup>, and Peter Muennig<sup>5</sup>

<sup>1</sup>Department of Psychiatry, University of Rochester Medical Center, Rochester, New York, USA

<sup>2</sup>Department of Family Medicine, University of Rochester Medical Center, Center for Communication and Disparities Research, Rochester, New York, USA

<sup>3</sup>Department of Public Health Sciences, University of Rochester Medical Center, Rochester, New York, USA

<sup>4</sup>Department of Society, Human Development, and Health, Harvard University School of Public Health, Boston, Massachusetts, USA

<sup>5</sup>Department of Health Management and Policy, Columbia University, Mailman School of Public Health, New York, New York, USA

### Abstract

**Background**—IQ is thought to explain social gradients in mortality. IQ scores are based roughly equally on Verbal IQ (VIQ) and Performance IQ tests. VIQ tests, however, are suspected to confound true verbal ability with socioeconomic status (SES), raising the possibility that associations between SES and IQ scores might be overestimated. We examined, first, whether two of the most common types of VIQ tests exhibited differential item functioning (DIF) favouring persons of higher SES and/or majority race/ethnicity. Second, we assessed what impact, if any, this had on estimates of the extent to which VIQ explains social gradients in mortality.

**Methods**—Data from the General Social Survey-National Death Index cohort, a US population representative dataset, was used. Item response theory models queried social-factor DIF on the Thorndike Verbal Intelligence Scale and Wechsler Adult Intelligence Scales, Revised Similarities test. Cox models examined mortality associations among SES and VIQ scores corrected and uncorrected for DIF.

---

**Correspondence to.** Dr Benjamin Chapman, Department of Psychiatry, University of Rochester Medical Center, 300 Crittenden, Rochester, NY 14620, USA; ben\_chapman@urmc.rochester.edu.

**Contributorship statement** BC, KF, PD, IK, PM: contributed to the conception and design of the work, and interpretation of data; contributed to drafting the work and critically revised it for important intellectual content; final approval of version to be published. BC, PM: contributed to the acquisition and analysis of data.

**Competing interests** None.

**Patient consent** No.

**Ethics approval** Institutional Review Board for the General Social Survey.

**Data sharing statement** The data is publicly available through the General Social Survey.

**Results**—When uncorrected for DIF, VIQ was correlated with income, education, occupational prestige and race, with correlation coefficients ranging between |0.12| and |0.43|. After correcting for DIF, correlations ranged from |0.06| to |0.16|. Uncorrected VIQ scores explained 11–40% of the Relative Index of Inequalities in mortality for social factors, while DIF-corrected scores explained 2–29%.

**Conclusions**—Two of the common forms of VIQ tests appear confound verbal intelligence with SES. Since these tests appear in most IQ batteries, circumspection may be warranted in estimating the amount of social inequalities in mortality attributable to IQ.

Socioeconomic status (SES) and cognitive ability are powerful predictors of health and longevity.<sup>1</sup> In fact, cognitive ability, as measured by the IQ, has been hypothesised to account for much of the SES-related health gradient.<sup>2</sup> Supporting this hypothesis, correlational studies suggest that those with stronger cognitive skills may be able to better understand medical instructions, navigate social bureaucracies, and avoid accidental death.<sup>3–8</sup> IQ also putatively correlates with markers of SES in the 0.4–0.55 range.<sup>9</sup>

IQ scores are themselves composed of two types of tests. ‘Performance IQ’ tests assess non-verbal reasoning and analytic ability. Tests of ‘Verbal IQ’ (VIQ) reflect language-based reasoning ability and knowledge.<sup>10</sup> ‘True’ VIQ is presumed to reflect factors, such as the speed and facility of language acquisition, and abstract and symbolic manipulation of language to solve problems and attain goals.<sup>10</sup> These skills assume different forms depending on the socioeconomic environment in which language is learned and reinforced, however.<sup>11,12</sup> Thus, language-based problem solving may vary markedly across socioeconomic strata.

In psychometric theory, VIQ ‘true scores’ refer to test scores putatively measuring the concept of ‘true VIQ.’<sup>13</sup> Under common statistical assumptions, VIQ true scores can be separated from other factors potentially influencing performance on VIQ tests that fall outside the definition of verbal intelligence. Many such factors have been implicated in VIQ test performance, including higher education, middle class SES and majority culture.<sup>11,12</sup> VIQ tests scores, therefore, run the risk of mixing or confounding verbal intelligence true scores with educationally or socially acquired knowledge, class-differential verbal styles, academic motivation, standardised testing experience and other residue of social position.

More technically, this measurement confounding arises from violations of collapsibility and exchangeability.<sup>14</sup> In other words, the association between the latent trait of ‘verbal intelligence’ (unobserved) and scores on a test of VIQ (observed) may not be collapsible across SES strata. Moreover, persons may not be exchangeable on non-intelligence attributes that affect response to items on the test (ie, measures of social standing). This type of systematic measurement error is called Differential Item Functioning (DIF).<sup>15</sup> If it is present to a significant degree on a VIQ test, persons of lower SES may achieve artificially deflated VIQ test scores, driving the apparent association between SES and VIQ upward.<sup>16</sup> As a result, the extent to which VIQ scores explain social gradients in mortality may appear larger than it actually is.

We examined whether two common forms of VIQ test exhibited DIF related to socioeconomic indicators (education, income, occupational prestige) and race in a US national sample. Although race/ethnicity may be less associated with social class in European countries, it is often considered a dimension of social stratification in the USA. We then compared SES correlations with VIQ scores, adjusted and unadjusted for DIF. Finally, we examined how much of the association between SES factors and all-cause mortality could be explained by VIQ scores, with and without correction for DIF. Our goal was to test a central premise of current research—that cognitive ability explains social patterns in mortality.

## METHODS

### Sample and design

We used data from the General Social Survey (GSS), an annual nationally representative sampling of US population social practices and attitudes. Conducted by the National Opinion Research Center at the University of Chicago, the GSS uses a multistage probability sampling of non-institutionalised adults age 18 years and over, with response rates from 70% to 82% in any given year,<sup>17</sup> yielding demographically identical annual samples. The GSS records age, gender, race/ethnicity, respondent occupation, income and years of education on the basis of face-to-face interviews with subjects. The Gallup-Thorndike Test of Verbal Intelligence<sup>18</sup> (hereafter Gallup-Thorndike) was administered during these interviews to one-third to one-half of the sample randomly selected during the years 1978, 1982, 1984, 1987–2000. We used data for 9381 persons with complete data for all variables of interest. Those lacking data (usually an SES indicator) were more likely female, younger, minority, and in worse self-rated health ( $p < 0.001$ ); the resulting analytic sample was still broadly similar to that of the USA in 2000.<sup>19</sup> The second VIQ test, the Wechsler Adult Intelligence Scales—Revised Similarities Test (WAIS-R Similarities), was administered in 1996 and yielded an analytic sample of 2444, by design demographically comparable to the broader Gallup Thorndike sample.

### Measures

Occupation was coded using the Socioeconomic Index (SEI), a continuous measure of occupational prestige, based on US Census information. Income was calibrated to 1990 US dollars. The Gallup-Thorndike<sup>18</sup> originally comprised 20 items taken from the Institute for Educational Research (IER) Intelligence Scale CAVD.<sup>20</sup> In the GSS, 10 of these items were administered in person by an interviewer. Tests of vocabulary are presumed to assess word familiarity, and also (1) concept formation (without which the correct definition cannot be given) and (2) the ability to deduce meaning of unfamiliar words based on known roots or syllables, using answer choices provided.<sup>21</sup>

The second VIQ test, the WAIS-R Similarities test,<sup>22</sup> presents persons with successively more difficult questions about how two different things are alike. For instance, an easy question might be ‘how are a fly and a mouse related?’ A completely correct response, such as ‘they are both animals’, receives two points. A response that is correct but does not

capture the similarity at the most abstract level (eg, ‘they both have eyes’) receives one point. The WAIS-R manual contains detailed guidelines for scoring responses as 0, 1, or 2.<sup>22</sup>

Vital status through 2008 was ascertained from the National Death Index. The validity of the National Death Index is typically high, with matching certainty arising from social security numbers and the additional identifiers in the GSS reaching 99.8%.<sup>23</sup> Further details on the GSS-National Death Index matching are available.<sup>17</sup>

## Analysis

Occupational prestige, income, and education were scaled by Relative Index of Inequalities (RII). The RII scores the person of highest standing on a social dimension as 0, and the lowest as 1.<sup>24,25</sup> A 1 unit change in regression models, therefore, is interpretable as a relative risk, but a specific kind: the risk at the absolute top, relative to the absolute bottom, of a distribution.

Item Response Theory (IRT) analyses of the Gallup-Thorndike were conducted with the Rasch model<sup>26</sup> and WAIS-R similarities analyses used the graded response model.<sup>27</sup> The Rasch model is formally equivalent to a mixed-effect logistic regression treating test items as repeated measures within person,<sup>28</sup> and estimating the probability of success or ‘difficulty parameter’ for each item *independently* of an examinee’s standing on the latent trait (random effect). The graded response model is an extension for ordered responses analogous to the extension from a binary to ordered logistic model (we relaxed the proportional odds assumption). The online supplementary material provides technical details of these IRT models and DIF analysis.<sup>29</sup>

Briefly, we examined DIF related to race, SEI, household income and education, as well as age and gender (which are correlated with SES) using interaction terms<sup>28</sup> in three increasingly stringent steps. In step 1, we screened social factor interaction terms separately for each item. In our second step, the model adjusted for all previously identified sources of DIF for a single item simultaneously. In the third stage, we adjusted for DIF factors across all items simultaneously. At each step, we retained those that were significant and met a DIF effect size threshold such that the item’s difficulty was 30% easier at one end of a sociodemographic dimension than at the other, irrespective of VIQ true score. Latent trait scores for each test were then estimated from IRT models with and without this final set of DIF interaction terms.

We examined impact of DIF on SES-VIQ associations via Pearson correlations between each SES factor and VIQ scores unadjusted and adjusted for DIF. We computed the absolute difference ( $r_{\text{unadj}} - r_{\text{adj}}$ ), as well as relative difference ( $r_{\text{unadj}}/r_{\text{adj}}$ ) between DIF adjusted and unadjusted score correlations.<sup>30</sup> We also estimated the association of DIF-corrected and uncorrected VIQ scores with mortality using Cox proportional hazards models with attained age as time scale and GSS baseline age as point of entry into the risk set,<sup>31,32</sup> fitting three models for each SES factor. Each model included gender as a covariate with time-varying hazards, based on preliminary proportionality analysis. The first model estimated the SES

---

Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jech-2013-203741>).

factor's RII, or the HR for those at the most disadvantaged, versus advantaged end of the distribution. A second model then added VIQ scores unadjusted for DIF and computed the excess hazard explained by these scores as  $(HR_{\text{unadjusted}} - HR_{\text{adjusted}}) / (HR_{\text{unadjusted}} - 1)$ , with 95% CIs obtained via bootstrap (1000 replicates). A third model then controlled for DIF-adjusted VIQ scores, again computing the change in estimate.

## RESULTS

Table 1 shows the sample demographics for the Gallup-Thorndike sample (left) and the WAIS-R similarities subset of that sample (right). With respect to VIQ tests, assumptions underlying IRT models appeared to be satisfactory.<sup>33</sup> For the Gallup-Thorndike, of the 40 interaction terms involving race, SEI, household income and education, 21 remained statistically significant and met the effect size criteria by the end of the three-stage screening. For the WAIS-R similarities test, of the 80 possible interactions, 14 remained significant and met the effect size threshold at the end of the third stage. Social factor DIF favoured white race and higher education, occupation and income. Additional age-related and gender-related DIF was observed on both tests, although the pattern did not consistently favour one gender or younger versus older persons. Online Supplementary table S1 lists the sources of DIF by item for each test. Social DIF seemed more apparent on vocabulary (9 out of 10 items) than similarities (5 out of 8 items).<sup>29</sup>

Table 2 reports the correlations between SES factors and VIQ test scores corrected and uncorrected for DIF. SES correlations with Gallup-Thorndike scores unadjusted for DIF were 0.16–0.33 larger in absolute magnitude, and 2.8–4.4 larger in relative magnitude, than unadjusted scores. For the WAIS-R Similarities, absolute differences in SES correlations ranged from 0.06 to 0.24 and relative differences from 2.0 to 2.6. DIF-adjusted correlations between SES factors and VIQ indicators fell outside the 95% CI of correlations with non-adjusted scores for all SES indicators on both tests.

Table 3 shows the RII as a HR for mortality for each social factor. Minority race exhibited non-proportional hazards (diminishing risk over the lifecourse), so estimates are presented at age 50 years. Table 3 also shows the change in estimate observed when controlling for latent trait scores adjusted and unadjusted for DIF. VIQ scores unadjusted for DIF accounted for smaller, but non-zero portions of social inequalities in mortality. For the Gallup-Thorndike, the change in estimate arising from corrected scores fell outside of the CI of that for uncorrected scores across three of four social factors. For the WAIS-R similarities test, the same pattern arose, but with wider CIs. Table 4 shows the RII for the WAIS-R similarities and Gallup-Thorndike scores with and without SES-corrected DIF. DIF-corrected scores showed smaller RIIs, with no appreciable difference in the proportion explained by SES. The latter quantity evidenced a very wide CI encompassing 0 in all cases. Sensitivity analyses revealed linearity in the log hazard for all factors, no VIQ social factor interactions or proportionality violations, nearly identical results excluding deaths within the first year, and comparable results with 2-parameter IRT models.

## DISCUSSION

Across two VIQ tests, we found DIF favouring persons of higher SES and/or majority race/ethnicity group. Correcting for this, DIF reduced correlations between VIQ scores and educational attainment, occupational status, income, as well VIQ differences between African-Americans and Caucasians. In turn, VIQ scores adjusted for DIF explained smaller amounts of social inequalities in mortality.

Some have argued that intelligence, rather than SES, is the fundamental cause of differentials in mortality.<sup>2</sup> This assertion is supported by many findings that cognitive ability test scores are substantial confounders of SES mortality risk.<sup>34</sup> Our findings suggest that DIF-corrected VIQ scores had slightly less association with SES and mortality than uncorrected ones, so a portion of the predictive power of VIQ may arise from indirect SES variance captured by VIQ test scores.

It is important to note, however, that small social differentials in VIQ still existed even when DIF was controlled. Accordingly, DIF-corrected VIQ scores continue to explain a modest portion of social gradients in mortality. This would suggest that VIQ is somehow involved in social inequalities in mortality, albeit to a smaller extent than has been presumed.

Environmental exposures<sup>35</sup> and early malnutrition<sup>36</sup> have documented effects on brain development and cognitive ability, and it is plausible, if not likely, that persons scoring lower on IQ tests, consequently, are challenged with respect to school performance, occupational advancement and earnings.<sup>37</sup> Thus, our data indeed suggest a legitimate—and probably reciprocal—association between VIQ and SES. Given the importance of this issue for policy, the critical question is not whether there is a link, but exactly how much measurement inaccuracy inflates our current estimates.

Specifically, DIF observed here may be explained by numerous factors affecting IQ test performance that are associated with social disadvantage. These include achievement motivation,<sup>38</sup> greater test performance anxiety and stress,<sup>3940</sup> fear that poor test scores will be used to perpetuate stereotypes about class and intelligence,<sup>4142</sup> lack of familiarity with test content among participants from lower SES and/or racial/ethnic minority subcultures,<sup>11</sup> different norms for, or uncertainty in, approaching test problems,<sup>4344</sup> use of different dialects, distrust of examiners administering the test,<sup>4546</sup> less familiarity with testing,<sup>47</sup> a lower reading level,<sup>4849</sup> and poorer test-taking skills.<sup>50</sup> The difficulty of disentangling verbal intelligence from factors relating to culture and academic achievement has been reported for some time.<sup>10</sup> Nevertheless, the fairness of IQ tests across SES is justified, in part, by reports that IQ tests with unknown degrees of SES DIF predict SES outcomes.<sup>37</sup> Such justifications may require reconsideration if VIQ tests confound Verbal IQ and SES to a non-trivial degree.

Our results must be interpreted with a balanced understanding of strengths and limitations. First, while these considerations suggest that VIQ tests capture educational and other SES variance, an important parallel argument has been offered: years of education, perhaps the most common index of SES, might actually measure some form(s) of intelligence, because cognitive abilities are generally required to achieve higher levels of education.<sup>51</sup> From this

viewpoint, adjusting any type of VIQ score for education-related DIF corrects for an IQ proxy and, thus, is an overcorrection. However, since ‘years of education’ is not a multi-item test score, IRT analyses cannot examine the issue. One future solution may be to use multi-item tests of academic achievement as a measure of education amenable to traditional IRT approaches and, thus, potentially separable from various forms of IQ. Second, VIQ is just one of two components of general IQ scores. Tests of the other component, Performance IQ, have been suggested by many,<sup>5253</sup> but not all,<sup>1112</sup> to avoid mixing SES with cognitive ability measurement. In this regard, many in the cognitive epidemiology community have begun to focus on measures of Performance IQ, including tests of reaction time or processing speed, as the key cognitive abilities predictive of mortality.<sup>8</sup>

Third, we only examine two common tests of VIQ. Although we did not study other tests, these two tests correlate highly with other VIQ tests (ie, 0.7 to 0.8),<sup>1054</sup> and with general IQ scores.<sup>1021</sup> Thus, we suspect that other VIQ tests, and general IQ scores, may be susceptible to some extent to this phenomenon. However, these results may or may not generalise to noncognitive psychological tests, such as personality measures, which may also be vulnerable to DIF and deserve study in their own right. It is also important to remember that SES is multidimensional, that some dimensions of SES might be more vulnerable to DIF than others, that different dimensions of SES may have differential associations with mortality, and these associations may vary at different points in the lifespan.

Although our analysis addresses these concerns for three common indicators of SES measured once, the use of other indicators would be helpful, such as the quality of education received or family social position. Longitudinal studies could examine the extent to which cognitive abilities at various points in the lifespan mediate prior SES-related health risks. Performance IQ, and/or tests based on theories of multiple intelligences,<sup>55</sup> may contribute better to our understanding of the inter-relationships between class, intelligence and health.

Ultimately, most IQ batteries used in epidemiologic study include vocabulary and/or similarities in VIQ tests. Thus, the behaviour of these tests will be transmitted to general or composite IQ scores, upon which many conclusions are based. If the other tests in the battery do not evidence social-factor DIF, overestimation of IQ-SES associations will be more attenuated. However, the number of other tests in the battery exhibiting similar DIF will dictate the degree of overestimation, and this is an unknown. Our findings thus constitute a ‘proof of principle’ suggesting care in interpreting data on IQ and social gradients in mortality.

## Acknowledgments

**Funding** This work was supported by US National Institutes of Health grants RC2MD004768, R01AG044588, and K08AG031328.

## REFERENCES

1. Roberts BW, Kuncel N, Shiner RN, et al. The power of personality: a comparative analysis of the predictive validity of personality traits, SES, and IQ. *Perspect Psychol Sci.* 2007; 4:313–346. [PubMed: 26151971]



2. Gottfredson LS. Intelligence: is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *J Pers Soc Psychol.* 2004; 86:174–199. [PubMed: 14717635]
3. Batty GD, Gale CR, Tynelius P, et al. IQ in early adulthood, socioeconomic position, and unintentional injury mortality by middle age: a cohort study of more than 1 million Swedish men. *Am J Epidemiol.* 2009; 169:606–615. [PubMed: 19147741]
4. Singh-Manoux A, Ferrie JE, Lynch JW, et al. The role of cognitive ability (intelligence) in explaining the association between socioeconomic position and health: evidence from the Whitehall II prospective cohort study. *Am J Epidemiol.* 2005; 161:831–839. [PubMed: 15840615]
5. Calvin CM, Deary IJ, Fenton C, et al. Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. *Int J Epidemiol.* 2011; 40:626–644. [PubMed: 21037248]
6. Batty GD, Deary IJ, Schoon I, et al. Mental ability across childhood in relation to risk factors for premature mortality in adult life: the 1970 British Cohort Study. *J Epidemiol Community Health.* 2007; 61:997–1003. [PubMed: 17933959]
7. Batty GD, Deary IJ, Macintyre S. Childhood IQ in relation to risk factors for premature mortality in middle-aged persons: the Aberdeen Children of the 1950s study. *J Epidemiol Community Health.* 2007; 61:241–247. [PubMed: 17325403]
8. Deary IJ, Weiss AW, Batty GD. Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychol Sci Public Interests.* 2010; 11:53–79.
9. Neisser U, Boodoo G, Bouchard TJ, et al. Intelligence: knowns and unknowns. *Am Psychol.* 1996; 51:77–101.
10. ASK. Assessing adolescent and adult intelligence. 3rd. New York: John Wiley & Sons; 2006.
11. Helms J. Why is there no study of cultural equivalence in standardized cognitive ability testing? *Am Psychol.* 1992; 47:1083–1101.
12. Helms, J. Contemporary Intellectual Assessment: Theories, Tests, and Issues. US: Guilford Press; 1998. The triple quandry of race, culture, and social class in standardized cognitive ability testing; p. 517-532.
13. Nunnally, JC.; Bernstein, IH. Psychometric Theory. 3rd. New York: McGraw-Hill, Inc.; 1994.
14. Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. *Epidemiol Perspect Innov.* 2009; 6:4. [PubMed: 19732410]
15. Millsap RE, Everson HY. Methodology review: statistical approaches for assessing measurement bias. *Applied Psychol Meas.* 1993; 17:297–334.
16. Rothman, KJ.; Greenland, S.; Lash, TL. Modern epidemiology. 3rd. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
17. Muennig P, Johnson G, Kim J, et al. The general social survey-national death index: an innovative new dataset for the social sciences. *BMC Res Notes.* 2011; 4:385. [PubMed: 21978529]
18. Thorndike RL. Two screening tests of verbal intelligence. *J Appl Psychol.* 1942; 26:128–135.
19. Bureau of the Census. Statistical Abstracts of the United States: 2001. Washington, DC: US Bureau of the Census; 2001.
20. Research IoE. I.E.R. Graded test of word knowledge. New York City: Teachers College, Columbia; 1923.
21. Jensen AR. Vocabulary and general intelligence. *Behav Brain Sci.* 2001; 24:1109–1110.
22. Wechsler, D. WAIS-R Manual: Wechsler adult intelligence scale—revised. San Antonio, TX: Psychological Corporation; 1981.
23. Hermansen SW, Leitzmann MF, Schatzkin A. The impact on national death index ascertainment of limiting submissions to social security administration death master file matches in epidemiologic studies of mortality. *Am J Epidemiol.* 2009; 169:901–908. [PubMed: 19251755]
24. Sergeant JC, Firth D. Relative index of inequality: definition, estimation, and inference. *Biostatistics.* 2006; 7:213–224. [PubMed: 16192414]
25. Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med.* 1997; 44:757–771. [PubMed: 9080560]

26. Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.
27. van der Linden, W.; Hambleton, RK. Handbook of modern item response theory. New York: Springer-Verlag; 1997.
28. Rijmen F, Tuerlinckx F, De Boeck P, et al. A nonlinear mixed model framework for item response theory. *Psychol Methods*. 2003; 8:185–205. [PubMed: 12924814]
29. Chapman, BP.; Fiscella, K.; Duberstein, PR., et al. Item Response Theory Technical Supplement for “Measurement Confounding Affects the Extent to Which Verbal IQ Explains Social Gradients in Mortality”. Vol. 5. Rochester, NY: University of Rochester Medical Center; 2014.
30. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health*. 1989; 79:340–349. [PubMed: 2916724]
31. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am J Epidemiol*. 1997; 145:72–80. [PubMed: 8982025]
32. Thiebaut AC, Benichou J. Choice of time-scale in Cox’s model analysis of epidemiologic cohort data: a simulation study. *Stat Med*. 2004; 23:3803–3820. [PubMed: 15580597]
33. Junker BW, Sijtsma K. Latent and manifest monotonicity in item response models. *Appl Psychol Meas*. 2000; 24:65–81.
34. Batty GD, Deary T. Education and mortality: the role of intelligence. *Lancet*. 2005; 365:1765–1766. [PubMed: 15910942]
35. McMichael AJ, Baghurst PA, Vimpani GV, et al. Tooth lead levels and IQ in school-age children: the Port Pirie Cohort Study. *Am J Epidemiol*. 1994; 140:489–499. [PubMed: 8067342]
36. Yang S, Platt RW, Kramer MS. Variation in child cognitive ability by week of gestation among healthy term births. *Am J Epidemiol*. 2010; 171:399–406. [PubMed: 20080810]
37. Nisbett RE, Aronson J, Blair C, et al. Intelligence: New findings and theoretical developments. *Am Psychol*. 2012; 67:130–159. [PubMed: 22233090]
38. Duckworth AL, Quinn PD, Lynam DR, et al. Role of test motivation in intelligence testing. *Proc Natl Acad Sci USA*. 2011; 108:7716–7720. [PubMed: 21518867]
39. Gass CS, Curiel RE. Test Anxiety in relation to measures of cognitive and intellectual functioning. *Arch Clin Neuropsychol*. 2011; 26:396–404. [PubMed: 21636602]
40. Allison DE. Test anxiety, stress, and intelligence-test performance. *Can J Behav Sci*. 1970; 2:26–37.
41. Steele CM, Aronson J. Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol*. 1995; 69:797–811. [PubMed: 7473032]
42. Croizet, JC.; Milet, M. Social class and test performance: from stereotype threat to symbolic violence and vice versa. In: Inzlicht, M.; Schmader, T., editors. *Stereotype threat: theory, process, and application*. New York: Oxford University Press; 2012. p. 118-201.
43. Olive H. Relationship of divergent thinking to intelligence, social class, and achievement in high-school students. *J Genet Psychol*. 1972; 121:179–186. [PubMed: 4650422]
44. Edwards TB. Deliberative attitudes, IQ, and social class. *J Exp Educ*. 1968; 37:7–18.
45. Klein RE, Freeman HE, Spring B, et al. Cognitive test-performance and indigenous conceptions of intelligence. *J Psychol*. 1976; 93:273–279.
46. Samuel W, Soto D, Parks M, et al. Motivation, race, social-class, and IQ. *J Educ Psychol*. 1976; 68:273–285.
47. Morgan AA, Marsiske M, Whitfield KE. Characterizing and explaining differences in cognitive test performance between african american and european american older adults. *Exp Aging Res*. 2008; 34:80–100. [PubMed: 18189169]
48. Manly JJ, Jacobs DM, Touradji P, et al. Reading level attenuates differences in neuropsychological test performance between African American and White elders. *J Int Neuropsychol Soc*. 2002; 8:341–348. [PubMed: 11939693]
49. Riley CS. Relationship between reading ability and verbal intelligence test performance. *Br J Educ Psychol*. 1966; 36:117.

50. Jones RN, Gallo JJ. Education and sex differences in the mini-mental state examination: effects of differential item functioning. *J Gerontol B Psychol Sci Soc Sci.* 2002; 57:P548–P558. [PubMed: 12426438]
51. Deary IJ, Johnson W. Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *Int J Epidemiol.* 2010; 39:1362–1369. [PubMed: 20504860]
52. Shuttleworth-Edwards AB, Kemp RD, Rust AL, et al. Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *J Clin Exp Neuropsychol.* 2004; 26:903–920. [PubMed: 15742541]
53. American Educational Research Association APA. Standards for Educational and Psychological Testing. Washington, DC: American Psychological Association; 1999. National Council on Measurement in Education.
54. Hunt, E. *Human Intelligence.* New York: Cambridge University Press; 2011.
55. Sternberg RJ, Castejón J, Prieto M, et al. Confirmatory factor analysis of the Sternberg Triarchic Abilities Test in three international samples: An empirical test of the triarchic theory of intelligence. *Eur J Psychol Assess.* 2001; 17:1.

**What is already known on this subject**

- ▶ General IQ scores are thought to partially explain social gradients in mortality.
- ▶ General IQ scores are composed of Performance IQ, and Verbal IQ (VIQ) tests.
- ▶ VIQ tests are suspected to confound true cognitive ability with socioeconomic status (SES).
- ▶ Measurement error may lead to overestimates of the extent to IQ explains social inequalities in mortality.

**What this study adds**

- ▶ Two common types of VIQ tests exhibit differential item functioning favouring persons of higher SES in a nationally representative US cohort.
- ▶ High correlations between VIQ scores and SES are inflated due to differential item functioning.
- ▶ Correction for differential item functioning reduces the explanatory role of IQ in social inequalities in mortality.

**Table 1**

Demographic composition of the analytic sample: 1978–2002 General Social Survey inked to the 2008 mortality via the National Death Index

	Mean/number	SD/percent of total sample
Age (years)	44.2	16.6
Gender		
Male	6422	45.1%
Female	7817	54.9%
Race/ethnicity		
White	11 894	83.5%
Black	1838	12.9%
Other race	507	3.6%
Socioeconomic status		
Occupational prestige <sup>*</sup>	42.15	13.9
Income <sup>†</sup>	44, 455	43 916
Education (years)	13.0	3.0

<sup>\*</sup>The occupational Prestige score reflects the 1970 Socioeconomic Index for 1979–1989, and the 1980 Socioeconomic Index for 1990–2002 respondents.

<sup>†</sup>Income is in year 2000 US dollars.

**Table 2**

Pearson correlations between SES, race, and verbal IQ test scores: 1978–2002 general social survey linked to the 2008 mortality via the National Death Index

	Latent trait scores, no DIF adjustment	Latent trait scores, DIF adjustment	Absolute difference	Relative difference
Gallup-Thorndike Verbal Intelligence test				
Occupational prestige	0.32 (0.30 to 0.34)	0.11 (0.09 to 0.13)	0.21	2.91
Education	0.43 (0.41 to 0.45)	0.10 (0.08 to 0.12)	0.33	4.30
Income	0.25 (0.23 to 0.27)	0.09 (0.07 to 0.11)	0.14	2.78
Minority race/ethnicity	−0.21 (−0.19 to −0.23)	−0.05 (−0.03 to −0.07)	0.16	4.20
Wechsler Adult Intelligence Scales Revised Similarities test				
Occupational Prestige	0.26 (0.23 to 0.32)	0.10 (0.06 to 0.14)	0.16	2.60
Education	0.40 (0.37 to 0.43)	0.16 (0.12 to 0.20)	0.24	2.50
Income	0.22 (0.18 to 0.26)	0.09 (0.05 to 0.13)	0.13	2.44
Minority race / ethnicity	−0.12 (−0.08 to −0.16)	−0.06 (−0.02 to −0.10)	0.06	2.00

Notes: Pearson correlations ( $r$ ) with bootstrap 95% CIs. For minority race/ethnicity, positive correlations indicate higher VIQ scores among minorities, negative correlations indicate lower VIQ scores. Income in Year 2000 US dollars. Occupational Prestige is 1970 Socioeconomic Index for 1979–1989, 1980 SEI for 1990–2002 respondents.  $n=9381$  for Gallup-Thorndike and 2444 for Similarities test.

DIF, Differential Item Functioning; VIQ, verbal IQ; SES, socioeconomic status.

**Table 3**

Social inequalities in mortality explained by biased and unbiased VIQ test scores

	<b>HR</b>	<b>Excess risk explained by VIQ scores uncorrected for DIF</b>	<b>Excess risk explained by VIQ scores corrected for DIF</b>
SES factors adjusted for Gallup-Thorndike test			
Occupational prestige	1.66 (1.34 to 2.05)	0.29 (0.11 to 0.47)	0.07 (0.01 to 0.13)
Education	1.92 (1.47 to 2.52)	0.40 (0.14 to 0.66)	0.29 (0.11 to 0.48)
Income	2.05 (1.71 to 2.45)	0.15 (0.05 to 0.24)	0.04 (0.00 to 0.07)
Minority Race/ethnicity	1.74 (1.52 to 1.98)	0.11 (0.05 to 0.17)	0.02 (0.00 to 0.03)
SES factors adjusted for WAIS-R Similarities Test			
Socioeconomic Index	1.94 (1.16 to 3.21)	0.13 (−0.27 to 0.72)	0.04 (−0.17 to 0.32)
Education	1.65 (0.90 to 3.01)	0.36 (−2.98 to 4.70)	0.09 (−1.30 to 1.52)
Income	2.34 (1.56 to 3.50)	0.04 (−0.23 to 0.27)	0.01 (−0.14 to 0.15)
Minority	1.44 (1.01 to 2.06)	0.09 (−0.26 to 0.59)	0.02 (−0.09 to 0.21)

Note. VIQ=verbal IQ. The HR is the Relative Index of Inequalities, or ratio of risk for those at the bottom of each distribution (ie, worst scoring or most disadvantaged) compared to those at the top. Percent of RII explained=excess risk explained by adjustment factor, computed as  $(HR_{unadjusted} - HR_{adjusted}) / (HR_{unadjusted} - 1)$  with 95% CIs based on bootstrap percentiles over 1000 replicates of 1000 observations. Same n=9381 for Gallup Thorndike and 2444 for WAIS-R similarities test. Results from separate Cox proportional hazard models for each factor, with age as time scale, all adjusted for gender as time-varying covariate due to non-proportional hazard. Minority status is also time varying, with estimates presented at age 50.



**Table 4**

## VIQ Inequalities in Mortality Explained by SES Indicators

	VIQ tests adjusted for SES Factors	
	Hazard ratio	Excess risk Explained by SES indicators
Gallup-Thorndike, uncorrected	1.49 (1.28 to 1.75)	0.73 (−0.05 to 1.02)
Gallup-Thorndike, DIF corrected	1.39 (1.20 to 1.61)	0.69 (−0.11 to 0.97)
Similarities, uncorrected	1.30 (0.94 to 1.79)	0.76 (−5.37 to 6.51)
Similarities, DIF corrected	1.17 (0.83 to 1.64)	0.79 (−6.2 to 9.73)

VIQ=verbal IQ. The HR is the Relative Index of Inequalities, or ratio of risk for those at the bottom of each distribution (ie, worst scoring or most disadvantaged) compared to those at the top. Percent of RII explained=excess risk explained by adjustment factor, computed as  $(HR_{unadjusted} - HR_{adjusted}) / (HR_{unadjusted} - 1)$  with 95% CIs based on bootstrap percentiles over 1000 replicates of 1000 observations. Same  $n=9381$  for Gallup Thorndike and 2444 for WAIS-R similarities test. Results from separate Cox proportional hazard models for each factor, with age as time scale, all adjusted for gender as time-varying covariate due to non-proportional hazard. Minority status is also time varying, with estimates presented at age 50 years. DIF, differential item functioning; SES, socioeconomic status.