

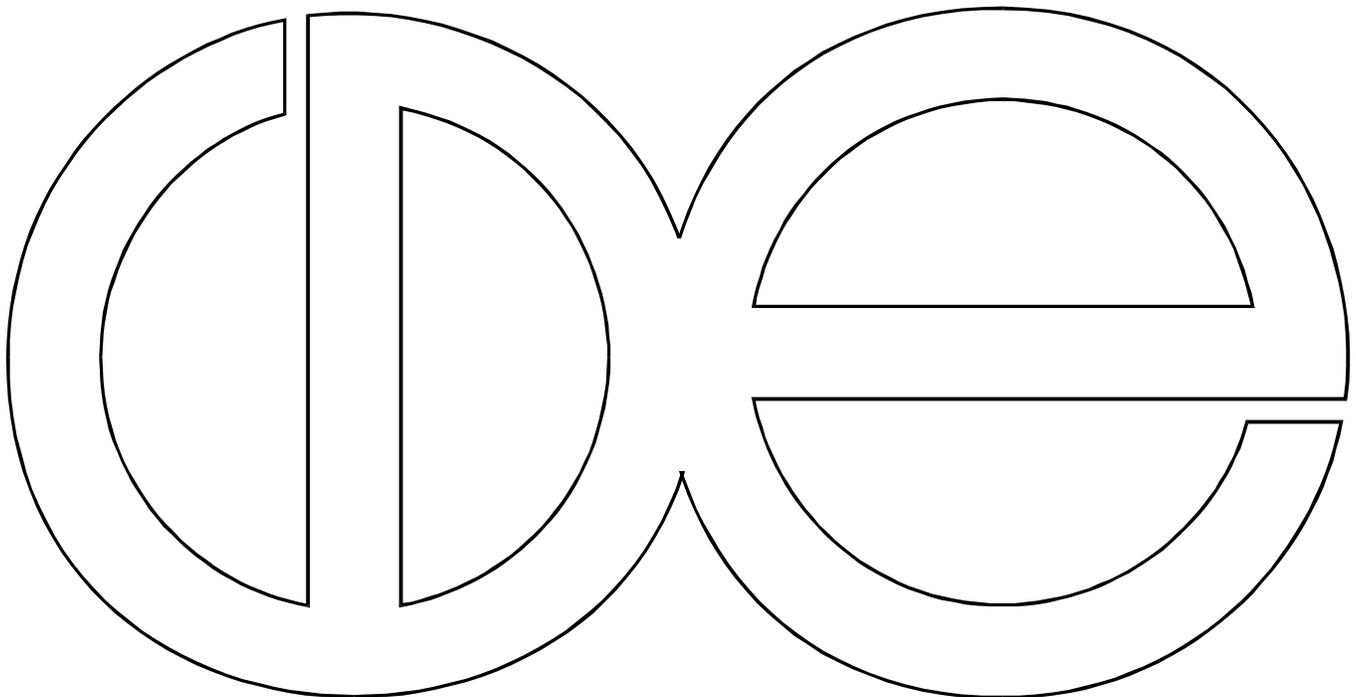
**Center for Demography and Ecology
University of Wisconsin-Madison**

**Trends in Black-White Test Score Differentials:
II. The WORDSUM Vocabulary Test**

Min-Hsiung Huang

Robert M. Hauser

CDE Working Paper No. 96-30



Trends in Black-White Test Score Differentials:

II. The WORDSUM Vocabulary Test

Min-Hsiung Huang

Robert M. Hauser

Department of Sociology
Center for Demography and Ecology
The University of Wisconsin-Madison

Rev. November 23, 1996

Prepared for the conference, "Intelligence on the Rise: Secular Changes in IQ and Related Measures," Emory University, April 1996. Support for this research was provided by the National Science Foundation (SBR-9320660); the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services; the Vilas Estate Trust; and the Center for Demography and Ecology at the University of Wisconsin-Madison. Data from the General Social Survey are available from the Interuniversity Consortium for Political and Social Research, University of Michigan, Ann Arbor, Michigan, or at <http://www.icpsr.umich.edu/gss/>. The opinions expressed herein are those of the authors.

There are a great many studies of the cognitive test performance of Blacks and Whites, but few can be used with confidence to measure differential performance trends (Hauser, 1996). One key resource, the series on Black and White school-children from the National Assessment of Educational Progress, begins only around 1970. Another series of data on verbal ability, which also began in the 1970s, provides clues about much longer trends in Black-White differences. These data are from the General Social Survey (GSS) of the National Opinion Research Center (NORC), which has regularly administered a 10-item verbal ability test (WORDSUM) to adult household members of all ages since the early 1970s (Davis & Smith, 1994). Because there are repeated, usually annual administrations of WORDSUM to cross-section samples of about 1000, it is possible to identify age and cohort effects on performance by assuming that there are no period effects. Using the GSS data, we can estimate trends in verbal ability among Blacks and Whites who were born from 1909 to 1974. These data show a consistent pattern of convergence between the test scores of Blacks and Whites throughout the period covered by the GSS cohorts. The convergence is not fully explained either by changes in social background or by changes in educational attainment across cohorts.

The ten GSS vocabulary items were chosen from "Form A," one of two parallel, twenty-item vocabulary tests selected by Thorndike. Each form contained two vocabulary test items from each of the levels of the vocabulary section of the Institute for Educational Research Intelligence Scale: Completion, Arithmetic Problems, Vocabulary, and Directions (Thorndike, 1942). Form A was developed by Thorndike in response to the need for a very brief test of intelligence that could be administered in social surveys (Thorndike & Gallup, 1944), and it was also used in an attempt to study the feasibility of an aptitude census (Thorndike & Hagen, 1952). Form A was later used by Miner (1957) in his monograph, *Intelligence in the United States*,

which attempted to assess the intellectual ability of the U.S. population using a national household sample. Alwin (1991) used the GSS WORDSUM data from 1974 to 1990 in an analysis which demonstrated that changes in family configuration could not account for the decline of verbal ability in the SAT-V.

For each of ten WORDSUM items, GSS respondents are asked to choose the one word out of five possible matches that comes closest in meaning to the word in capital letters. Figure 1 gives a set of sample items that are similar to those in WORDSUM.¹ The GSS obtains personal interviews, and each item is handed to the respondent on a preprinted card. Before 1988, WORDSUM was administered to the full GSS sample, but only every other year. Since 1988 it has been administered to two-thirds of the sample in each survey year, using an alternate forms design. From 1974 to 1994, WORDSUM was completed by 11,160 Whites and 1,418 Blacks who were ages 20 to 65 at the survey date and who also provided valid data on years of schooling, number of adults, number of siblings, and structure of the family of orientation.

Miner (1957, p. 28-30) argued that vocabulary tests are highly correlated with tests of general intelligence. He assembled some thirty-six studies in which a vocabulary measure had been correlated with a measure of general intelligence, and he found a median correlation of 0.83. Miner also noted that, whereas Wechsler had originally excluded a vocabulary test from the Wechsler-Bellevue scales because “he felt it might be unfair to illiterates and those with a foreign language background,” he later decided that it was an excellent measure of general intelligence. Finally, Miner noted that the median correlation of verbal tests with full-scale IQ tests is larger than the median correlation Wechsler reported between the Wechsler-Bellevue Full Scale and fifteen other measures of general intelligence (Wechsler, 1944). More recently, Wolfle (1980, p.

110) reported that the correlation between the GSS vocabulary test and the Army General Classification Test (AGCT) was 0.71.

Despite these indications of validity, we would urge caution in the use of WORDSUM. First, the test is very short, so its reliability is low. The internal consistency reliabilities are 0.712 among Whites and 0.628 among Blacks. In the 1994 GSS, in addition to WORDSUM, half the sample was administered 8 of the 14 similarity (abstract reasoning) items from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981).² The correlations between WORDSUM and the WAIS-R similarity scores were 0.394 for Whites and 0.409 for Blacks, without correction for unreliability. After correction for internal consistency reliability, the correlations were 0.589 among Whites and 0.608 among Blacks. Second, as we will discuss in more detail, despite the common use of vocabulary tests in IQ instruments and the high correlations between tests of vocabulary and of general intelligence, there is evidence of divergent trends between IQ and verbal ability in the past several decades. Thus, even if verbal ability is a valid indicator of general intelligence, neither the overall trend in WORDSUM nor the specific trends in WORDSUM among Blacks and Whites need follow those in more general tests of ability.

Figure 2 shows the vocabulary test score distributions of Blacks and Whites in the GSS from 1974 to 1994. The mean numbers of correct items are $\bar{x}_w = 6.23$ ($s_w = 2.09$) for Whites and $\bar{x}_b = 4.78$ ($s_b = 1.93$) for Blacks. From the display of relative frequencies, it is obvious that there is a ceiling effect on the scores for Whites in the general population. Similarly, there are ceiling or floor effects on the number of correct answers in other subpopulations, e.g., persons with many or few years of schooling. For this reason, our analyses of the WORDSUM data are based on a

two-sided Tobit specification, which compensates for censoring at both ends of the distribution on the assumption that the true distribution of test scores is Gaussian (Maddala, 1983).

Methodological issues

Before attempting a trend analysis of the GSS data, we carried out several methodological analyses. Only 460 GSS respondents (3.7 percent) refused to answer any of the WORDSUM items, and we ignored them throughout the analysis. The refusals had completed slightly fewer years of schooling ($\bar{x} = 10.5$ vs. 12.9 years), but were also more variable in schooling ($s = 3.6$ vs. 2.8 years). We considered whether item nonresponses -- other than complete refusal -- should be treated as incorrect answers, or whether total scores should be adjusted for the number of items answered. We compared the internal consistency reliabilities among complete responses and among incomplete responses, treating nonresponses in the latter group as errors. These were virtually the same, 0.62 and 0.64. Further, the correlations between educational attainment and WORDSUM among the complete and incomplete responses were identical. Thus, we assumed that item-specific nonresponses were erroneous responses.

We also experimented with a range of corrections for guessing by assuming that respondents who missed easy items may have answered harder items correctly by guessing. Under each guessing scenario, we recoded the harder "correct" answers as erroneous and examined the correlations of WORDSUM with educational attainment, occupational status (Stevens & Featherman, 1981; Stevens & Cho, 1985), and -- for 1994 GSS respondents only -- the WAIS-R similarity items. We found that the correlations with external criteria were lowered by corrections for guessing, so we did not introduce such a correction.

Finally, we carried out logistic regression analyses of each item, by race and by birth cohort, in order to look for differential item functioning (DIF). We carried out each analysis with and without adjustment of total test scores for unreliability, and in each case we controlled age at test administration. In the absence of differential item functioning, the odds of a correct response to an item would be the same for two groups at every ability level, as indicated by the total test score. Uniform differential item functioning occurs when the odds of answering an item are equally greater or equally lower at every ability level. Nonuniform differential item functioning occurs when the difference between groups in the odds of a correct answer are not the same across ability levels (Swaminathan and Rogers, 1990).

In looking for differential item functioning, we modified the usual model for differential item functioning by including dummy variables for age at testing as covariates. For Whites, we based the intercohort comparisons on 10-year birth cohorts from 1910-19 to 1960-69. Among Blacks, we collapsed the 1910-29 birth cohort in order to increase sample size. The usual logistic regression model for predicting the probability of a correct response to an item is:

$$P(u = 1) = e^z / (1 + e^z), \quad (1)$$

where

$$z = \tau_0 + \tau_1 \theta + \tau_{2j} w_j + \tau_{3j} (\theta \times w_j). \quad (2)$$

The variable w_j indicates membership in the j th group, and θ is the observed ability of an individual. In this case, θ is the respondent's total WORDSUM score, w_j is the respondent's birth cohort, and $\theta \times w_j$ is the product of the two independent variables, w_j and θ . An item shows uniform DIF if $\tau_{2j} \neq 0$ for some j and $\tau_{3j} = 0$ for all j , and it shows nonuniform DIF if $\tau_{3j} \neq 0$ for

any j , whether or not $\tau_{2j} = 0$. Since WORDSUM varies with age at testing, we also included dummy variables for respondent's age in the regression. Thus, the model becomes:

$$z = \tau_0 + \tau_1 \theta + \tau_{2j} w_j + \tau_{3j} (\theta \times w_j) + \tau_{4i} x_i, \quad (3)$$

where x_i is a dummy variable for membership in the i th 10-year age group. We estimated this model separately for Whites and Blacks, and we looked for evidence that items had become easier or harder for successive cohorts. We used conventional statistical tests with $p = 0.01$, supplemented by BIC, the Bayesian information criterion (Raftery, 1995), and we also looked for systematic variation across cohorts in the parameters for differential item functioning (DIF). By all standards, we found little evidence of nonuniform DIF, and in only one case, a relatively difficult item among Whites, did the Bayesian criterion indicate there was reliable evidence of uniform differential item functioning. However, there were four items among Whites and two items among Blacks in which there was nominally significant uniform DIF. Among Blacks, both items had become successively more difficult, and among Whites, two items had become more difficult, and two had become less difficult.

We also estimated a model similar to equation 3, in which we pooled the data for Blacks and Whites and specified effects of racial-ethnic group, corresponding to those of cohorts in equation 3. In comparing Blacks and Whites, when total WORDSUM scores are controlled, a few items are slightly easier for Whites than for Blacks with the same WORDSUM score. After adjustment for internal consistency reliability, those effects are reversed: Five items are significantly easier for Blacks than for Whites with the same estimated true scores. One other item is then significantly easier for low-scoring Blacks and significantly harder for high-scoring Blacks. There does not appear to be any relationship between item difficulty and differential item

functioning. Given these equivocal findings, it would probably be a good idea to test the sensitivity of our regression analyses to the elimination of selected items. Because of the small number of items in WORDSUM, we have not done so here. However, these findings provide yet more reason for caution in the interpretation of WORDSUM trends.

Finally, we looked for external evidence that the difficulty of the WORDSUM items may have changed across time. Between 1921 and 1967, four studies have reported frequency counts and ranks for English words (Thorndike 1921, 1931; Thorndike & Lorge, 1944; Kučera & Francis, 1967). Unfortunately, the several sets of rankings are of uncertain comparability. Thorndike (1921) is a list of the 10,000 most frequent words “in a count of about 625,000 words from literature for children; about 3,000,000 words from the Bible and English classics; about 300,000 words from elementary-school text books; about 50,000 words from books about cooking, sewing, farming, the trades, and the like; about 90,000 words from the daily newspapers; and about 500,000 words from correspondence” (p. iii). Thorndike (1931) reports the addition of “extensive additional counts from over 200 other sources including about 5,000,000 words” (p. iii), and he extended the list to 20,000 words. Thorndike and Lorge (1944) added information from counts of an additional 4.5 million words. Kučera and Francis (1967) did not build on the work of Thorndike and Lorge, but analyzed “a collection of statistical information obtained from analysis of The Standard Corpus of Present-Day Edited American English, a computer processible corpus of language texts assembled at Brown University during 1963-64” (p. xvii). The corpus of more than 1,000,000 words was selected from the press (reportage, editorial, and reviews), religion, skills and hobbies, popular lore, belles lettres and biography, learned and scientific writings, humor, and several categories of fiction writings. It would appear that the Thorndike-

Lorge series was cumulative, beginning with a focus on children's reading material and later extending to more general collections of text. For that reason, that series is not entirely appropriate for an assessment of trend. On the other hand, the Kučera-Nelson corpus is almost entirely made up of adult-oriented text, and it is thus not strictly comparable to the Thorndike-Lorge series.

For each WORDSUM item, we looked up or estimated the approximate rank of each stimulus word, of its synonym, and -- in most cases -- of a plausible distractor. We had no basis for interpreting changes in the relative frequency of key words within a given item. Kučera and Francis (1967) identified the 50,000 most common words, and in many cases, words that had appeared in the earlier lists were much less common in the 1967 list. We do not know whether differences between rankings in that list and the earlier lists are due to true temporal change or to other differences in the selection of text. In any event, we find that the ranks of WORDSUM items have been either stable or decreasing. That is, if frequency of usage in written text is an indicator of difficulty, WORDSUM has become somewhat more difficult across time, independent of any other change in verbal ability in the general population. That tendency occurs to some degree across the lists of 1921, 1931, and 1944, and it appears strongly when we include the 1967 list in the comparison. However, we do not believe that changes in item difficulty, if they are real, could account for decreasing differences between the test scores of Blacks and Whites.

Trend analyses of WORDSUM

We estimate three models of intercohort trends in verbal ability in the 1974 to 1994 GSS data. In each case, we analyze the total number of correct answers to the ten WORDSUM items using a two-sided Tobit specification. The estimated coefficients and their standard errors are

shown in Table 1. The baseline model includes sex, race, age, birth cohort, and interactions between race and sex and between race and birth cohort:³

$$E[y] = \alpha + \sum_1^I \beta_i x_i + \sum_1^J \gamma_j w_j + \delta_1 z_1 + \delta_2 z_2 + \delta_3 (z_1 z_2) + \sum_1^J \lambda_j z_1 w_j, \quad (4)$$

where y is the number of correct WORDSUM items, α is the intercept, the x_i are dummy variables for age groups, the β_i are age effects, the w_j are dummy variables for birth cohorts, the γ_j are cohort effects, z_1 is a dummy variable for race, z_2 is a dummy variable for sex, the δ s are effects of sex and race, and the λ_j are effects of race by cohort interactions.

The model is intended to describe trends in verbal ability among Blacks and Whites, free of the confounding influences of temporal changes in the age composition of the population and of the association of age with test scores. If we had observations from a single cross-section sample, we could not conceivably separate the effects of chronological ages from those of birth cohorts. Because the GSS provides repeated cross-section measures in samples of the same cohorts at different ages, we can estimate distinct effects of chronological age and year of birth. However, the model requires a strong identifying assumption, namely, that there are no period effects on test scores. That is, we assume that there are no effects on test scores, specific to the year of the survey, above and beyond the combination of age and birth cohort effects that pertain to persons in each survey year.

In the case of one item, we were particularly concerned about the assumption of no period effects because a large corporation began using one of the stimulus words in the name of a product line in the late 1970s. We would have expected this change of usage to increase

knowledge of the meaning of the stimulus word. For this reason, we carried out a separate analysis of trends in knowledge of this word, in which we estimated period and age effects, but not cohort effects. We found a possible trend toward increasing item difficulty, but it began before and ended shortly after initial the commercial use of the word. Thus, we do not think the commercial use of the word has affected its validity for our purposes.

As shown in Table 1, effects have been defined relative to black women aged 20 to 29 who were born in 1900 to 1919. In the oldest cohort, Whites answered almost 2.5 more items correctly than Blacks. There was almost no overall gender difference in verbal scores, and White men scored insignificantly lower than the combined effects of race and gender would predict. In the baseline model, persons aged 30 to 39 and 40 to 49 each obtained scores about 0.50 words higher than persons aged 20 to 29. Persons aged 50 to 59 had only about half the advantage of 30 to 49 year olds, relative to those at ages 20 to 29, and persons aged 60 to 65 had a small and statistically insignificant advantage over the youngest age group. There is an irregular upward progression in the coefficients for birth years (γ_j) and an irregular downward progression in the effects of interactions between birth years and race (λ_j). These patterns imply that verbal ability scores have been increasing among Blacks and decreasing among Whites.

The social background model adds eight social background characteristics to the baseline model of equation 4: Father's educational attainment, mother's educational attainment, father's occupational status, number of siblings, non-intact family (at age 16), foreign residence (at age 16), farm background, and Southern residence (at age 16). In addition, the social background model includes three dummy variables that flag missing values on father's education, mother's education, and father's occupation. The education model adds respondent's years of schooling to

the social background model. Years of education are coded into a series of twenty-one dummy variables with twelve years of education as the reference group. Comparisons among estimates from these models may help us understand the effects of social background and respondent's years of education on Black-White differences and their trends over time. That is, changes across models in the effects of birth cohorts among Blacks and Whites tell us the degree to which the trends can be explained by the make-up of Black and White cohorts in terms of social background and educational attainment.

The effects of age in the three models are also of interest in their own right. They have been estimated with 20 to 29-year-olds as the reference group. In the baseline model and, again, when social background has been controlled, test scores increase by about half a point at ages 30 to 39 and 40 to 49 and by about half that amount at ages 50 to 59 and 60 to 65. However, this age pattern is partly a consequence of the variation in schooling by cohort in the GSS data. When years of school are also controlled, there remains a smaller increase in test scores at ages 30 to 39 and 40 to 49, about 0.3 years relative to persons aged 20 to 29. However, the two older age groups no longer have any advantage relative to the youngest age group. This age pattern would appear to contradict some suggestions in the research literature that a relatively crystallized ability, like vocabulary knowledge, increases indefinitely with age.

Figure 3 shows the intercohort trends in WORDSUM scores of Blacks and Whites born from 1909 to 1974, as estimated in the baseline model. Black-White differences in test scores have been decreasing.⁴ In the most recent birth cohort, 1970 to 1974, the Black-White difference is 16.4 percent as large as in the earliest birth cohort, 1909 to 1919.⁵ The differences decline irregularly across birth cohorts, but never increase substantially from one cohort to the next.

Relative to the oldest cohort, they are 84 percent, 63 percent, 64 percent, 64 percent, 56 percent, and 16 percent as large in successive cohorts.⁶ About two-thirds of the convergence can be attributed to an upward trend in test scores among Blacks, from 4.0 in the oldest cohort to 5.4 in the youngest. About one-third of the convergence can be attributed to a downward trend among whites, from 6.5 in the oldest cohort to 5.8 in the youngest. Most of the White decline took place in cohorts born after 1950, while the growth in Black test scores took place between the cohorts of 1909 and 1949 or after the 1960s.

Figure 4 shows the trends in test scores for Blacks and Whites when social background variables are added to the baseline model. Because there have been persistent but declining differences in social background between Black and White cohorts, one might expect the initial differences between Black and White test scores to be smaller and the convergence to be less rapid when social background is controlled. This appears to be the case.⁷ The initial difference between Blacks and Whites in the cohort of 1909 to 1919 is 1.7 points in the social background model, while it was 2.5 points in the baseline model. Moreover, convergence is greater than in the baseline model. Black-White differences in WORDSUM disappear in the 1970 to 1974 birth cohort when social background variables are controlled. The convergence appears to have been driven largely by a continuous decline in the verbal ability of Whites, while the test scores of Blacks varied irregularly from one cohort to the next. That is, the estimates suggest that verbal ability was relatively stable among Blacks with the same social background throughout most of this century, while it declined steadily among Whites of constant social background throughout the century. Black-White differences in social background account for part of the initial

difference between average test scores, and improvements in the social background of Blacks relative to Whites account for part of the convergence.

Figure 5 shows the intercohort trends in test scores when years of schooling as well as social background are added to the baseline model.⁸ Net of schooling, as one would expect, average Black and White test scores are less far apart. In the oldest cohorts, the differential is more than one correct answer on the WORDSUM test (half a standard deviation), and in all but the youngest cohort, the differential is slightly less than one correct answer. Thus, much of the Black-White difference in vocabulary scores is associated with differences in social background and schooling. They account for more than a 1-point test score difference in the oldest cohort and for slightly less than a half-point difference in the next-to-youngest cohort.

The data also suggest that changes in social background and schooling contributed to the convergence in test scores. For example, in the baseline model, the change in the Black-White difference between the oldest and youngest cohorts was 2.1 points; in the social background model, it was 1.8 points; and in the schooling model, it was 1.2 points. Moreover, there are also signs of convergence in test scores, above and beyond the salutary effects of social background and schooling. That is, in the schooling model, the Black and White test score series are almost parallel, but there are additional signs of convergence in the youngest cohort. Finally, among Blacks and Whites, the net trends in Figure 5 are close to linear. Both groups show an overall decline in vocabulary knowledge, excepting the relatively high performance of the youngest cohort of Blacks. Thus, as successive cohorts completed higher nominal levels of schooling, those levels of schooling may have become less selective with respect to a basic level of verbal

proficiency.⁹ Some might argue that this finding indicates a decline in school quality, but we prefer to reserve such terminology for direct evidence that the effects of schools have declined.

Social Background, Size of Sibship, and the Trend in Verbal Ability

After controlling social background, the trend in verbal ability is downward, both among Blacks and Whites. For this reason, it would be easy to ignore the positive effects of intercohort changes in social background on verbal ability. Table 2 shows mean levels of the eight social background variables used in the preceding analysis. There have been very large and, for the most part, unidirectional shifts in the composition of successive cohorts on six of the eight background variables.¹⁰ For this reason, we have also estimated the effect of changes in social background on verbal ability.

Average levels of maternal and paternal educational attainment have grown rapidly, by five years or more among Whites and by seven years or more among Blacks, even as the variability in parental schooling has declined. For example, the growth in schooling of the fathers of Blacks -- 7.7 years -- is 1.7 times its standard deviation in the cohort of 1909 to 1919 and 2.5 times its standard deviation in the cohort of 1970 to 1974. Among Whites, father's occupational status has grown by almost as much as its standard deviation in the oldest cohort; among Blacks, father's occupational status has grown by close to half its standard deviation in the older cohorts.

The number of siblings has declined rapidly, from 4.7 to 2.4 among Whites and from 6.0 to 4.2 among Blacks, and there has also been a decline in the variability of sibship size. However, relative to the standard deviations, the change has not been as large as in the case of parental schooling. Among Whites the decline in size of sibship is about 1.3 times as large as the standard deviation in the youngest cohort, and among Blacks the decline is less than half as large as the

standard deviation in the youngest cohort. Among Whites, the share of children who were not living with both own parents was about 20 percent for cohorts born before 1959, but began to grow rapidly thereafter. Among Blacks, the share of children raised in non-intact families was much higher than among Whites early in the century. It rose after the end of World War I to 47 percent, and after a modest decline in the Depression cohorts, it has grown steadily to cover almost two-thirds of Black adults.

In cohorts born from 1909 to 1974, very few adult Americans were living in a foreign country at age 16, although more were perhaps born abroad. There has been very little change in this share over time. On the other hand, the share of the population with rural origins has declined dramatically. The percentage with farm background has declined from 38 to 19 among Whites and from 58 to 10 among Blacks. Thus, Blacks in the youngest cohort were much less likely than Whites to be of rural origin. There has been little change among Whites in the share of adults, 26 to 30 percent, who lived in the South at age 16, but Southern origin declined rapidly among Blacks, from 79 percent in the cohort of 1909 to 1919 to 53 percent in the cohort of 1970 to 1974.

Figure 6 shows the effects of intercohort changes in social background on trend in WORDSUM among Blacks and Whites. That is, it displays the changes in WORDSUM that are implied by changes in social background alone, disregarding the actual trends net of social background, gender, and age at testing that were displayed in Figure 4. The estimates were constructed by applying the regression coefficients of the social background model of Table 1 to the means of the social origin variables in Table 2 (including effects of the dummy variables for missing data):

$$\hat{y}_{jg} = \hat{\alpha}_g + \sum_1^K \hat{\pi}_k \bar{s}_{jkg}, \quad (5)$$

where \hat{y}_{jg} is the predicted WORDSUM score of the j th cohort in racial-ethnic group g , $\hat{\alpha}_g$ is the intercept of group g , $\hat{\pi}_k$ is the estimated effect of the k th social background variable, and \bar{s}_{jkg} is the mean of the k th social background variable for the j th cohort in group g . The steeper pair of lines in Figure 6 shows the combined effects of intercohort changes in all eight social background variables among Blacks and Whites. The other two lines show the effects of intercohort changes in the observed mean number of siblings in each cohort. That is, the former series shows the combined effect of all changes in social background on WORDSUM, and the latter series isolates the effect of change in numbers of siblings. We have looked separately at the effects of change in size of sibship because of persistent scholarly interest in the effects of that variable on ability (Blake, 1989) and occasional efforts to connect changes in average ability with fertility decline (Zajonc, 1976; Zajonc, 1986; Alwin, 1991).

Intercohort changes in social background have had profound implications for observed levels of verbal ability. Among Whites, the implied growth in WORDSUM is 1.46 items from the oldest to the youngest cohorts, and among Blacks, the implied growth is 1.82 items. That is, improvements in social background across U.S. cohorts born between 1909 and 1974 imply growth of about 0.7 standard deviations in verbal ability among Whites and more than 0.9 standard deviations in verbal ability among Blacks. However, changes in numbers of siblings contribute very modestly to these trends. When changes in size of sibship are analyzed jointly

with those of other social background characteristics, declining sibship size accounts for only an increase of 0.21 correct WORDSUM items among Whites and 0.16 correct WORDSUM items among Blacks. These effects account for only 14 percent of the effects of social background on trend in WORDSUM among Whites and 9 percent of the effects of social background on trend in WORDSUM among Blacks. By contrast, the effects of change in parental schooling are much larger, 0.79 items among Whites and 1.09 items among Blacks. That is, growth in parental schooling accounts for more than half of the salutary effects of intercohort changes in social background on verbal ability.

Why does number of siblings contribute so little to the effects of social background on intercohort change in verbal ability? First, number of siblings is correlated with each of the other social background variables, and these other variables also affect WORDSUM. Thus, the effect of number of siblings is substantially less in the regression analysis than it might appear to be if other social background characteristics were ignored. Second, although the declining numbers of siblings in the GSS data reflect the sharp fertility decline of the first half of this century, there have also been very large and perhaps larger changes in other significant social background variables. Third, those other social background variables -- especially parental schooling -- are strongly associated with verbal ability.

Summary and Discussion

To the evidence that Black-White differences in academic achievement were reduced in cohorts born after the middle 1960s we can add new evidence of a longer term convergence between Blacks and Whites in verbal ability. Under reasonable assumptions with respect to the effects of age and sex, data collected by the NORC GSS since the middle 1970s suggest a 65-year

trend toward convergence between adult Blacks and Whites in verbal ability, as indicated by a short test of vocabulary knowledge. Differences between Blacks and Whites in social background and schooling account for a large share of the observed differences in test scores. The convergence in verbal ability appears to have been close to complete for persons of similar background and schooling who were born after 1970. Improvements in the social background and schooling of Blacks account in part for the convergence in test scores, but other factors also appear to have reduced the test-score difference, especially in recent cohorts.

In combination, the effects of social background on verbal ability and the positive intercohort changes in social background that have occurred throughout the 20th century imply substantial intercohort growth in verbal ability, almost a standard deviation among Blacks and almost three-quarters of a standard deviation among Whites. Thus, if we could generalize from effects on verbal ability to those on other abilities, we should count improvements in social background among the sources of long-term cognitive growth in the United States and, perhaps, in other nations. The effects of changes in social background are only to a small degree consequences of fertility decline. Rather, they are mainly a consequence of improvements in parental schooling, occupational standing, and geographic origin (farm background and Southern origin).

At the same time, a singular aspect of the GSS series is that they show a long-term decline in verbal ability in the White population. A similar trend appears for Blacks and Whites when social background and schooling have been controlled, though the decline among Blacks is less than that among Whites. Are the test-score series from the GSS plausible in light of Flynn's findings of rising IQ scores (Flynn, 1984; Flynn, 1987), both in the United States and in nineteen

other countries? We believe that the GSS series are not necessarily inconsistent with Flynn's findings, though we certainly hope that it will be possible to test those findings directly or indirectly with data from other sources. While the evidence is weak, our examination of word frequency ranks suggests that the difficulty of the WORDSUM items may have increased over the decades. Also, Loehlin, Lindzey, & Spuhler's review of global trends in intelligence took note of two studies in which verbal scores had declined across time, while other ability measures had increased (1975, p. 135- 39). It should also be recalled that, in Flynn's initial report of temporal gains in test scores, much of the discussion concerned the possible inconsistency between IQ gains up through 1978 and the post-1963 declines in the SAT-V (1984, p. 36-39). In his recent review, "IQ gains over time," Flynn notes that tests may be ordered from those with reduced cultural content, "many of which are indicators of fluid intelligence," to those that measure crystallized intelligence, with "less emphasis on-the-spot problem solving and more on whether someone has acquired the skills, or general knowledge, or vocabulary we would expect an intelligent person to gain in a normal life" (1994, p. 617). Thus, he places "pure vocabulary tests" at the extreme of tests of crystallized intelligence.

Flynn's reading of the available evidence, worldwide, is that

IQ gains over time diminish as tests get farther and farther from measuring fluid intelligence. ... Verbal IQ gains vary from almost nil to 20 points per generation, with 9 as a rough median, and some of this is adult data from military testing.

Among the eleven countries that allow a comparison, there is not one in which verbal gains match the gains on culture-reduced, or performance, or nonverbal tests and often the ratios run against verbal gains by two or three to one. Where

vocabulary gains can be distinguished from verbal gains in general, they rarely match them.

If we accept this reading, we should not be entirely surprised to find a valid time-series of vocabulary measures with a pronounced downward trend. At the same time, the clear distinction between the behavior of vocabulary tests and other tests of ability across time is a valuable reminder of the limits of the present findings.

In this context, leaving aside our specific findings in the GSS, we might want to consider other applications of similar research designs. There is no hope of going as far back in time in the measurement of fluid intelligence as the GSS series permits us in the assessment of verbal ability. However, we could cover a considerable span of cohorts by administering more comprehensive tests to a single large cross-section sample, provided only that we truly understood the intrinsic relationship between age and performance on each test. Alternatively, a new series of ability measures in repeated cross-section samples, following the GSS model, would permit estimation both of age and cohort effects and could provide both retrospective and prospective assessments of trends and differentials in abilities.

REFERENCES

- Alwin, D. F. (1991, October). Family of Origin and Cohort Differences in Verbal Ability. *American Sociological Review*, 56(5), 625-38.
- Blake, J. (1989). *Family Size and Achievement*. Berkeley: University of California Press.
- Davis, J. A., & Smith, T. W. (1994). *General Social Surveys, 1972-1994: Cumulative Codebook*. Chicago: National Opinion Research Center.
- Flynn, J. R. (1984). The Mean IQ of Americans: Massive Gains 1932 to 1978. *Psychological Bulletin*, 95(1), 29-51.
- Flynn, J. R. (1987). Massive IQ Gains in 14 Nations: What IQ Tests Really Measure. *Psychological Bulletin*, 101(2), 171-191.
- Flynn, J. R. (1994). IQ Gains Over Time. In R. J. Sternberg, S. J. Ceci, Horn, J. J. Matarazzo & S. Scarr (Eds.), *Encyclopedia of Human Intelligence* (pp. 617-623). New York: Macmillan.
- Hauser, R. M. (1996). Trends in Black-White Test Score Differentials: I. Uses and Misuses of NAEP/SAT Data. In U. Neisser (Ed.), *The Rising Curve: Long-term Gains in IQ and Related Measures*. Washington, D.C.: American Psychological Association.
- Hauser, R. M., & Featherman, D. L. (1976, April). Equality of Schooling: Trends and Prospects. *Sociology of Education*, 49, 99-120.
- Hauser, R. M., & Mare, R. D. (1993). A General Social Survey Mini-Module on Trends and Differentials in Cognition [A Proposal to the Governing Board of the GSS]. Center for Demography and Ecology, UW-Madison.

Kučera, H., & Francis, W.N. (1967). *Computational Analysis of Present-Day American English*. Providence, Rhode Island: Brown University Press.

Loehlin, J. C., Lindzey, G., & Spuhler, J. (1975). *Race Differences in Intelligence*. San Francisco: W.H. Freeman and Company.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Miner, J. B. (1957). *Intelligence in the United States: A Survey*. New York: Springer.

Raftery, A.E. (1995). Bayesian Model Selection in Social Research. In P.V. Marsden (Ed.), *Sociological Methodology 1995* (pp. 111-63). Cambridge: Blackwell.

Stevens, G., & Cho, J.H. (1985). Socioeconomic Indexes and the New 1980 Census Occupational Classification Scheme. *Social Science Research, 14*, 142-168.

Stevens, G., & Featherman, D.L. (1981). A Revised Socioeconomic Index of Occupational Status. *Social Science Research, 16*, 364-395.

Swaminathan, H., & Rogers, H. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement, 27*, 361-370.

Thorndike, E.L. (1921). *The Teacher's Word Book*. New York: Teachers College, Columbia University.

Thorndike, E.L. (1931). *A Teacher's Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People*. New York: Teachers College, Columbia University.

Thorndike, E.L., & Lorge, I. (1944) *The Teacher's Word Book of 30,000 Words*. New York: Teachers College, Columbia University.

Thorndike, R. L. (1942). Two Screening Tests of Verbal Intelligence. *Journal of Applied Psychology*, 26, 128-135.

Thorndike, R. L., & Gallup, G. H. (1944). Verbal Intelligence of the American Adult. *Journal of General Psychology*, 30, 75-85.

Thorndike, R. L., & Hagen, E. P. (1952). *Analysis of Results of Field Trials to Determine the Feasibility of an Aptitude Census*. Human Resources Research Center Research Bulletin, vol. 52-22.

Wechsler, D. (1944). *The Measurement of Adult Intelligence (3rd Ed.)*. Baltimore: Williams and Wilkins.

Wechsler, D. (1981). *WAIS-R Manual: Wechsler Adult Intelligence Scale-Revised*. San Antonio: The Psychological Corporation.

Wolfe, L. M. (1980, April). The Enduring Effects of Education on Verbal Skills. *Sociology of Education*, 53, 104-114.

Zajonc, R. B. (1976). Family Configuration and Intelligence. *Science*, 192, 227-36.

Zajonc, R. B. (1986). The Decline and Rise of Scholastic Aptitude Scores: A Prediction Derived from the Confluence Model. *American Psychologist*, 41, 862-867.

FOOTNOTES

1. We have selected these items at random from the list in Miner (1957, p. 53). The National Opinion Research Center has requested that we not reveal the specific items used in WORDSUM.
2. Hauser & Mare (1993) added the similarity items in order to calibrate them against WORDSUM for use in supplementary telephone interviews with brothers and sisters of the GSS respondents.
3. The number of cases for Whites born in 1909-19, 1920-29, 1930-39, 1940-49, 1950-59, 1960-69, and 1970-74 are 543, 1304, 1660, 2542, 2860, 1555, and 187, respectively. The counts for Blacks are 60, 186, 298, 426, 587, 342, and 28, respectively. The small number of cases in the oldest and youngest cohorts of Blacks implies that the findings about them should be interpreted with caution.
4. In the baseline model, the likelihood-ratio test statistics for intercohort differences in test scores and for the race by cohort interactions are both highly significant.
5. Again, because of the small number of cases in the youngest cohort, this observation should not be taken too seriously.
6. These estimates and others cited below refer specifically to differences between White and Black women. Because the race by gender interactions are negligible, the estimates are similar for differences between White and Black men.
7. In the social background model, the likelihood-ratio test statistics for intercohort differences in test scores and for the race by cohort interactions are both highly significant.
8. In the schooling model, the likelihood-ratio test statistic for intercohort differences in test scores is highly significant. However, the race by cohort interactions are only marginally

significant ($p < 0.03$); thus, the cohort trend lines for Blacks and Whites are almost parallel.

9. Also, recall that there may have been some increase in the difficulty of WORDSUM items, as indexed by word frequency.

10. See Hauser & Featherman (1976) for additional evidence of intercohort change in social background.

Table 1. Reductions in the Black-White Difference on the National Assessment of Educational Progress

	White-Black Difference, in Standard Deviations		
	1969-73	1990	Change
<i>9 year olds</i>			
Science	1.14	0.84	-0.30
Math	0.70	0.54	-0.16
Reading	0.88	0.70	-0.18
<i>Average</i>	0.91	0.69	-0.21
<i>13 year olds</i>			
Science	0.96	0.76	-0.20
Math	0.92	0.54	-0.38
Reading	0.78	0.40	-0.38
<i>Average</i>	0.89	0.57	-0.32
<i>17 year olds</i>			
Science	1.08	0.96	-0.12
Math	0.80	0.42	-0.38
Reading	1.04	0.60	-0.44
<i>Average</i>	0.97	0.66	-0.31
<i>Overall average</i>	0.92	0.64	-0.28

Source: Mullis, et al. (1991), as reported by Herrnstein & Murray (1994: 291), who stated, "The computations assume a standard deviation of 50."

Table 2. Means and Standard Deviations of Selected Social background Variables: General Social Survey

Social Background Variable	Birth Cohort						
	1909-1919	1920-1929	1930-1939	1940-1949	1950-1959	1960-1969	1970-1974
Whites							
Father's education	7.6 (4.1)	8.4 (4.0)	9.0 (3.9)	10.4 (3.8)	11.7 (3.8)	12.6 (3.6)	13.4 (3.2)
Mother's education	8.1 (3.8)	8.9 (3.6)	9.7 (3.4)	10.8 (3.1)	11.7 (3.0)	12.3 (2.8)	13.1 (2.6)
Father's occupational status	28.3 (15.4)	28.9 (15.7)	30.4 (16.7)	33.5 (18.5)	37.1 (19.8)	39.8 (20.4)	42.9 (21.6)
Number of siblings	4.7 (3.1)	4.1 (3.2)	3.9 (3.2)	3.5 (2.9)	3.5 (2.6)	3.2 (2.4)	2.4 (1.8)
Not living with both own parents at age 16	0.23 (0.42)	0.22 (0.41)	0.22 (0.41)	0.19 (0.39)	0.20 (0.40)	0.27 (0.44)	0.34 (0.47)
Lived in foreign country at age 16	0.02 (0.14)	0.04 (0.18)	0.05 (0.21)	0.03 (0.17)	0.03 (0.17)	0.03 (0.16)	0.01 (0.11)
Farm background	0.38 (0.49)	0.35 (0.48)	0.33 (0.47)	0.27 (0.44)	0.24 (0.42)	0.23 (0.42)	0.19 (0.39)
Lived in the South at age 16	0.27 (0.44)	0.26 (0.44)	0.28 (0.45)	0.29 (0.45)	0.27 (0.44)	0.30 (0.46)	0.30 (0.46)
Blacks							
Father's education	4.4 (4.4)	6.7 (4.1)	7.3 (3.8)	8.7 (4.3)	9.4 (3.8)	11.4 (3.5)	12.1 (2.9)
Mother's education	6.7 (4.4)	8.1 (4.3)	8.2 (3.6)	9.4 (3.5)	10.3 (3.4)	11.9 (3.0)	13.8 (2.6)
Father's occupational status	21.4 (9.0)	21.0 (8.4)	21.8 (8.2)	25.2 (15.2)	24.7 (13.2)	30.5 (18.5)	25.2 (10.2)
Number of siblings	6.0 (4.6)	5.3 (3.8)	6.0 (4.4)	5.4 (3.7)	5.7 (3.7)	5.0 (4.0)	4.2 (4.1)
Not living with both own parents at age 16	0.37 (0.48)	0.47 (0.50)	0.41 (0.49)	0.43 (0.50)	0.46 (0.50)	0.51 (0.50)	0.63 (0.49)
Lived in foreign country at age 16	0.00 (0.00)	0.02 (0.13)	0.05 (0.22)	0.06 (0.24)	0.07 (0.25)	0.04 (0.21)	0.01 (0.12)
Farm background	0.58 (0.50)	0.44 (0.50)	0.37 (0.48)	0.27 (0.44)	0.20 (0.40)	0.12 (0.33)	0.10 (0.31)
Lived in the South at age 16	0.79 (0.41)	0.73 (0.45)	0.69 (0.46)	0.57 (0.50)	0.52 (0.50)	0.49 (0.50)	0.53 (0.50)

Figure 1. Illustrative Vocabulary Test Items

a. LIFT	1. sort out	2. raise	3. value	4. enjoy	5. fancy
b. CONCERN	1. see clearly	2. engage	3. furnish	4. disturb	5. have to do with
c. BROADEN	1. efface	2. make level	3. elapse	4. embroider	5. widen
d. BLUNT	1. dull	2. drowsy	3. deaf	4. doubtful	5. ugly
e. ACCUSTOM	1. disappoint	2. customary	3. encounter	4. get used	5. business
f. CHIRrup	1. aspen	2. joyful	3. capsize	4. chirp	5. incite
g. EDIBLE	1. auspicious	2. eligible	3. fit to eat	4. sagacious	5. able to speak
h. CLOISTERED	1. miniature	2. bunched	3. arched	4. malady	5. secluded
i. TACTILITY	1. tangibility	2. grace	3. subtlety	4. extensibility	5. manageableness
j. SEDULOUS	1. muddled	2. sluggish	3. stupid	4. assiduous	5. corrupting

Source: Miner (1957: 53).

Figure 2. WORDSUM Distributions of Blacks and Whites:
General Social Survey, 1974-94

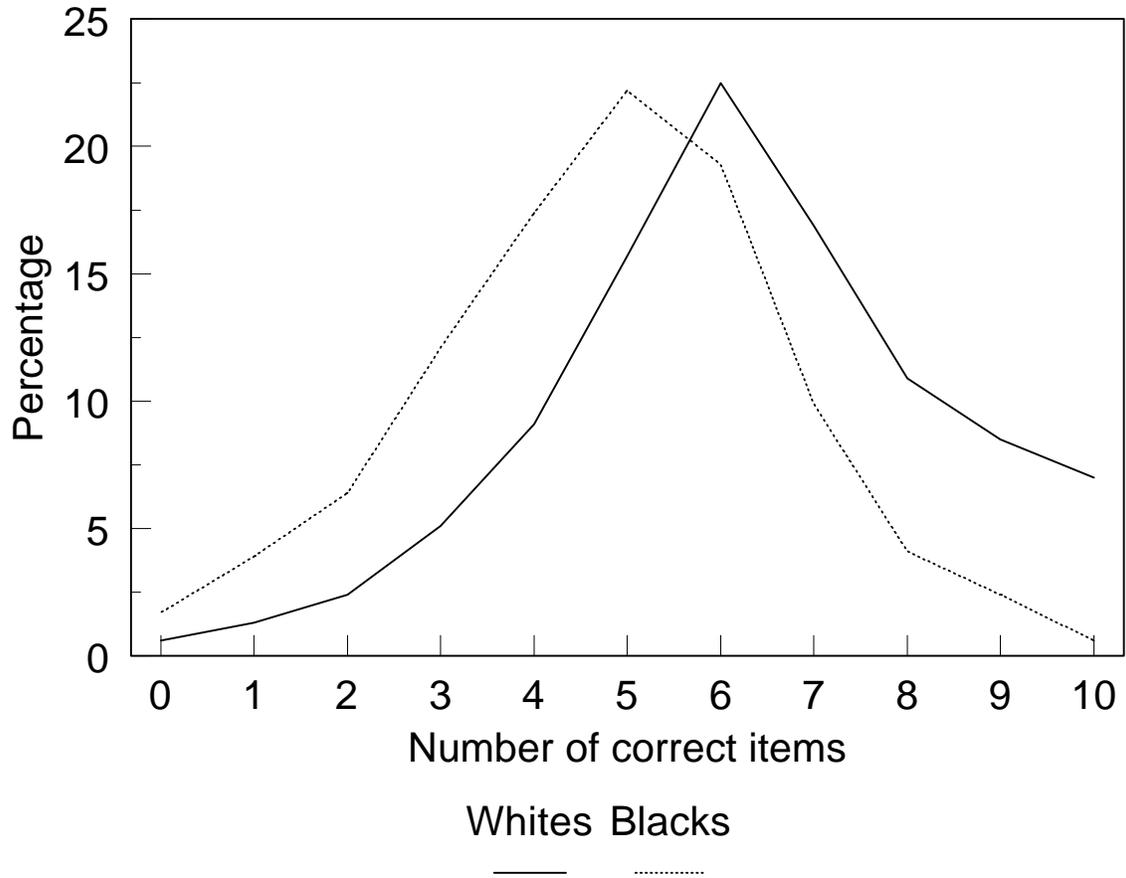


Figure 3. Baseline Model of WORDSUM Trends:
General Social Survey, 1974-94

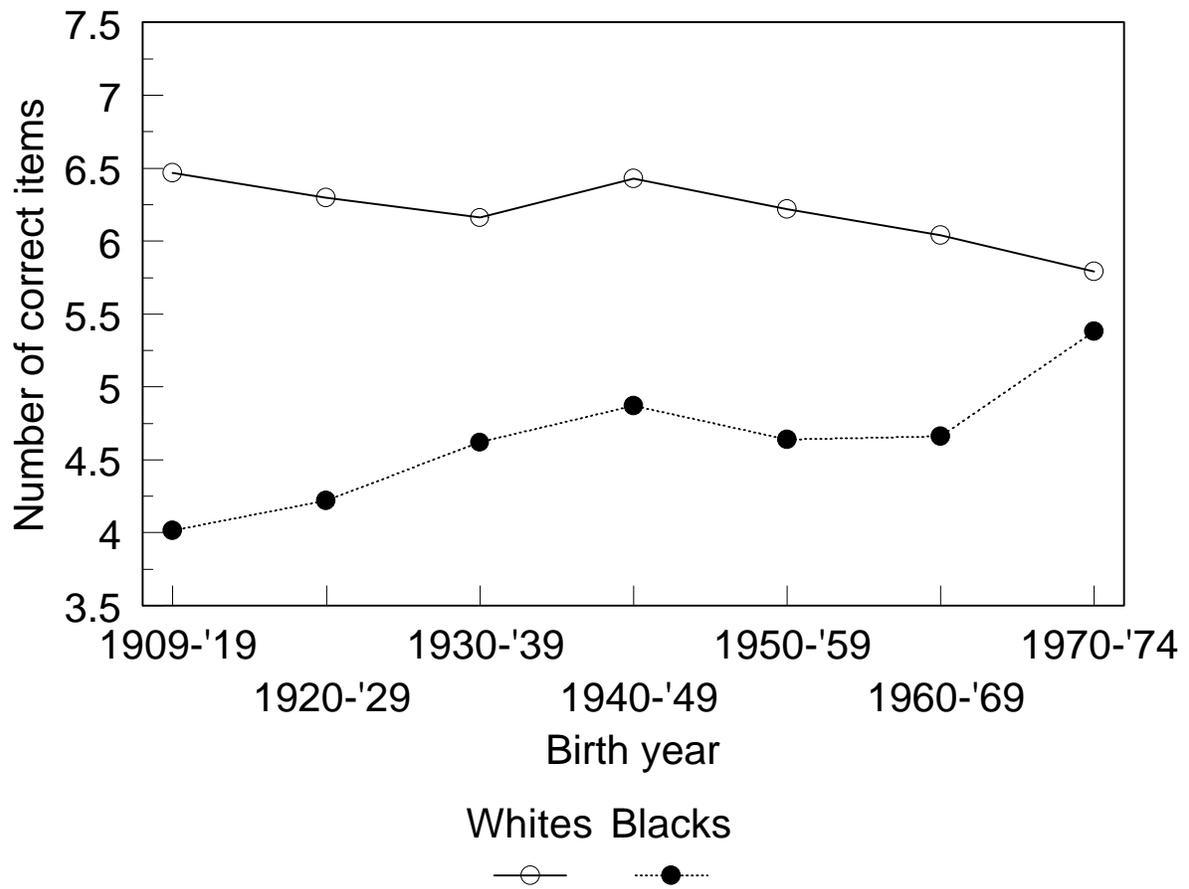


Figure 4. Social Background Model of WORDSUM Trends:
General Social Survey, 1974-94

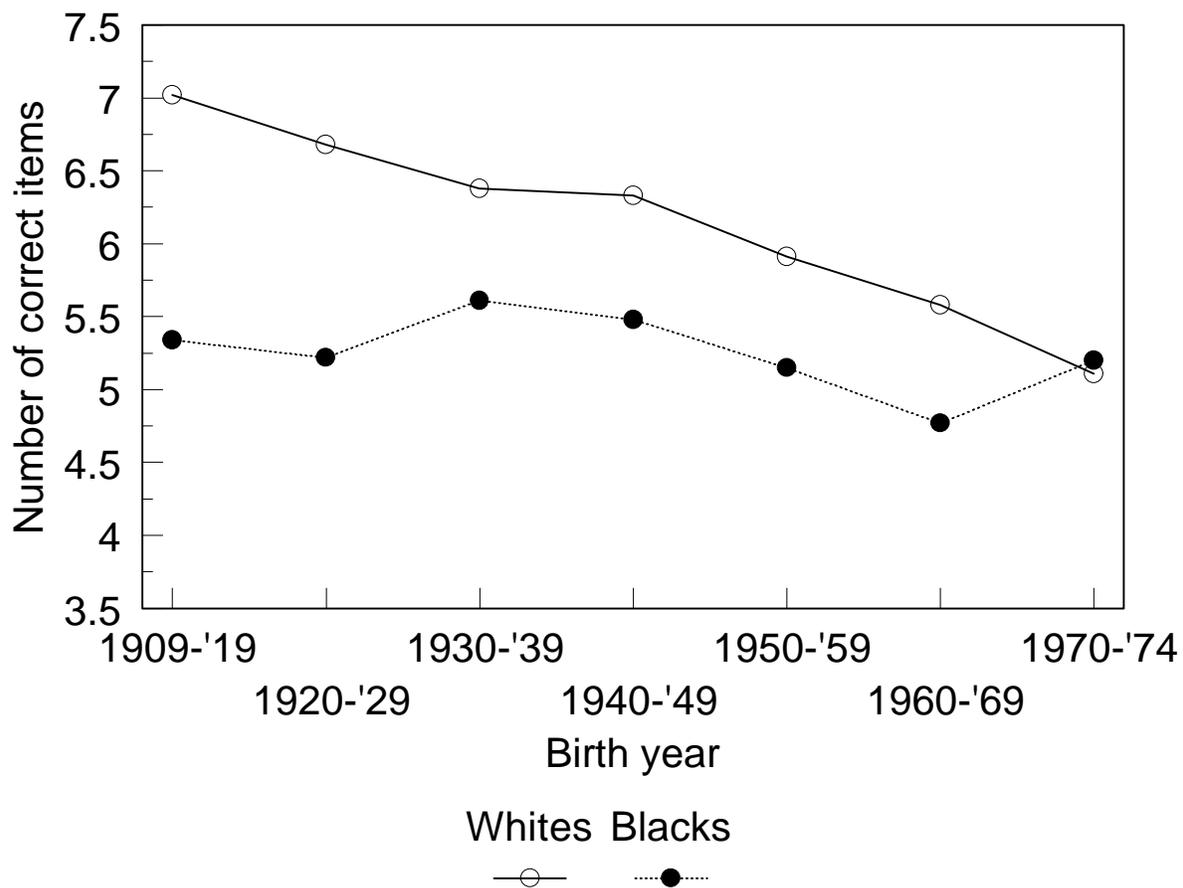


Figure 5. Education Model of WORDSUM Trends:
General Social Survey, 1974-94

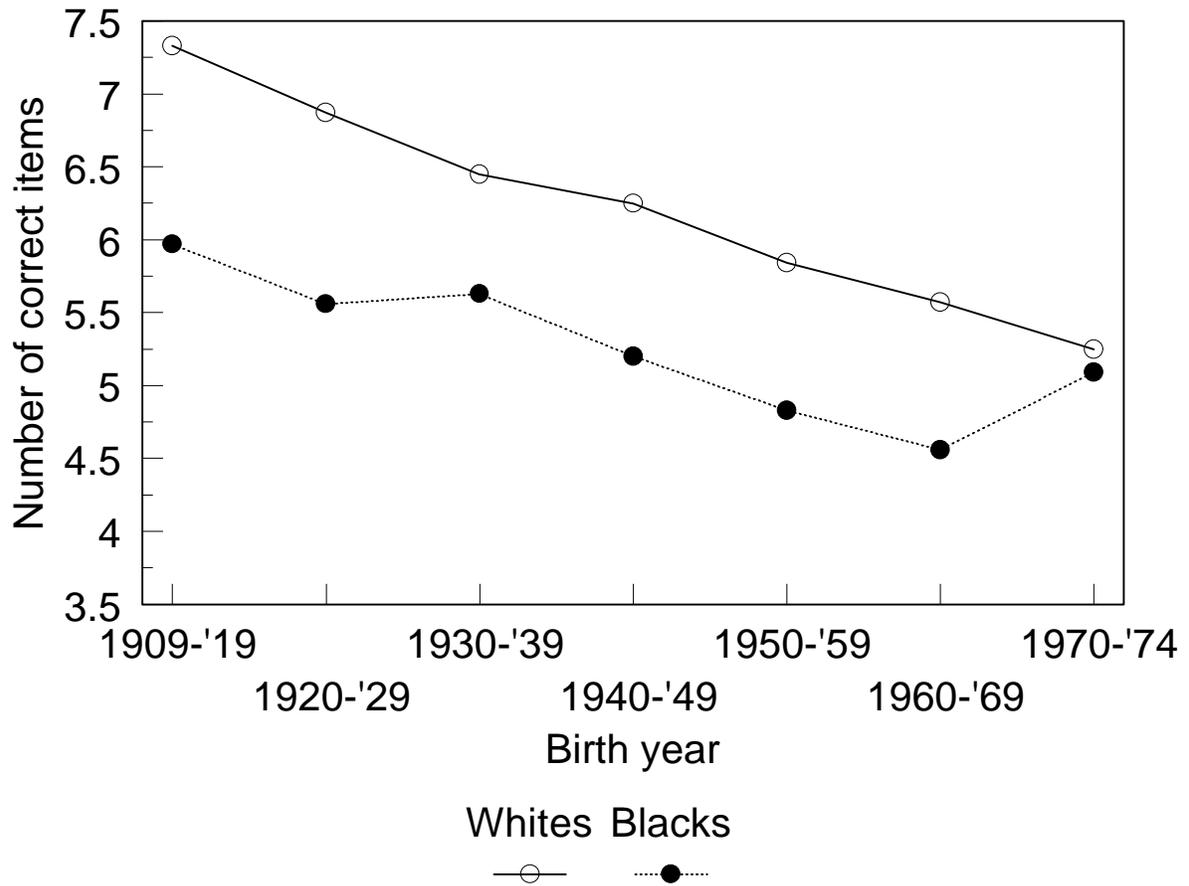
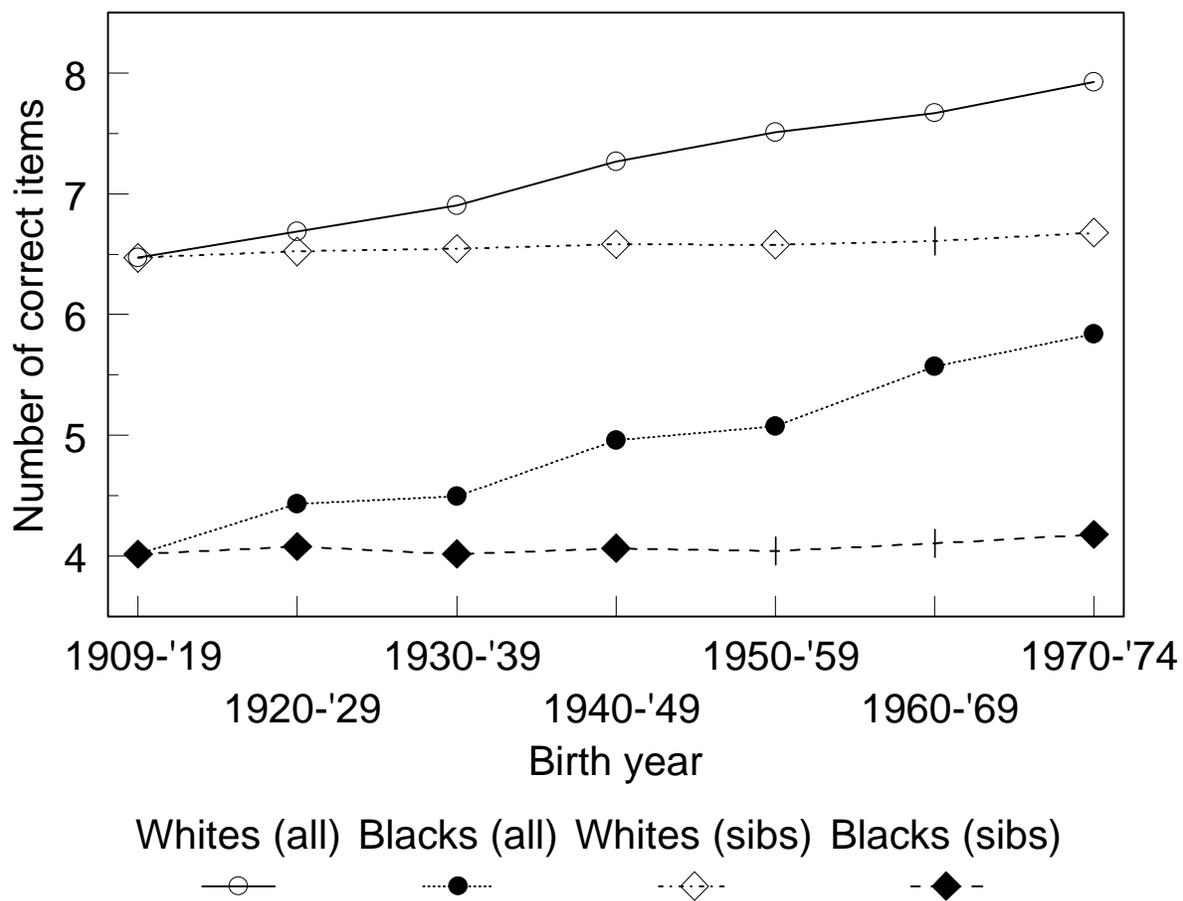


Figure 6. Effects of Intercohort Changes in Social Background on WORDSUM Trends: General Social Survey, 1974-94



Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
comments to: mhuang@ssc.wisc.edu
requests to: cdepubs@ssc.wisc.edu