EURAMERICA Vol. 31, No. 2 (June 2001), 185-230

Convergent Trends in Black-White Verbal Test-Score Differentials in the U.S.: Period and Cohort Perspectives*

Min-Hsiung Huang

Institute of European and American Studies
Academia Sinica

Robert M. Hauser

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

* Support for this research was provided by the Institute of European and American Studies, Academia Sinica, Taipei, Taiwan, and by 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, and the data files used in this analysis are available directly from the authors. We appreciate the comments of Tony Tam and our two anonymous reviewers. Address correspondence and reprint requests to Min-Hsiung Huang, Institute of European and American Studies, Academia Sinica, Taipei 11529, Taiwan, or e-mail to mhhuang@sinica.edu.tw.
Received February 7, 2001; accepted June 4, 2001
Proofreader: Yi Ping Chang

Abstract

We extend and modify our earlier intercohort analysis of General Social Survey (GSS) data through 1998 and confirm our previous finding of a very significant, long-term Black-White convergence, which is attributable to improvements in socioeconomic background and schooling among African-Americans. We maintain that cohort analysis is essential for an accurate understanding of temporal changes in the Black-White test score gap. Survey year analysis, on the other hand, is likely to obscure changes over time in Black-White test score differences because change occurs primarily through cohort replacement. However, even in an analysis of aggregate changes in the Black-White test score gap across survey years, when conducted properly, we find that the Black-White test score gap narrowed significantly over the period from 1974 to 1998.

Key Words: Black-White test score gap, cohort analysis, period analysis, verbal ability, WORDSUM.

I. Introduction

Has the Black-White gap in cognitive ability in the U.S. been persistent over time? Although there has been a controversy about what cognitive ability tests really measure, cognitive ability is usually measured by IQ tests and other similar tests, such as scholastic aptitude tests and school achievement tests. Using data from large-scale surveys with nationally representative samples, some researchers report that the Black-White gap in academic achievement test scores has narrowed over the past 30 years (Campbell, Hombo, and Mazzeo, 2000, Hedges and Nowell, 1999), but that the convergence has decreased or ceased after the late 1980's.[1] Huang and Hauser (1998) use data from the General Social Survey (GSS), 1974-1994, and report a 65-year intercohort

[1] The Black-White gap on IQ tests and other similar tests has typically been about one standard deviation (Herrnstein and Murray, 1994; Jensen, 1980; Neisser *et al*. 1996). If test scores are normally distributed, a gap of one standard deviation suggests that 84% of American Whites exceed the mean score among American Blacks and about 16% of American Blacks exceed the mean score among American Whites. This, however, is no longer the case on many cognitive tests (Hauser, 1998; Hedges and Nowell, 1999; Huang and Hauser, 1998; Jencks and Phillips, 1998). The National Assessment of Education Progress (NAEP) provides data from three decades of student performance on reading, science, and math for students ages 9, 13, and 17 from 1969 to 1999. Data from the NAEP suggest that the Black-White gap in reading was reduced by 20%, 26%, and 42% for students ages 9, 13, and 17, respectively, from 1971 to 1999. For math, these figures are 20%, 30%, and 23% for students ages 9, 13, and 17, respectively, from 1973 to 1999 (Campbell, Hombo, and Mazzeo, 2000). Hedges and Nowell (1999) use data from several national representative samples of adolescents from 1965 to 1996 and report that the Black-White test score gap in the composite and the reading test scores had been reduced by 30% from 1965 to 1992.

trend toward convergence between Black and White adults in verbal ability. Lynn (1998) also uses data from the GSS, but he suggests that the Black-White gap in verbal ability has been persistent over the period from 1974 to 1996. Some other researchers recognize the fact that a gap between test scores of Blacks and Whites has narrowed through the 1980's, but they suggest that the gap could begin to widen again (Herrnstein and Murray, 1994: 293).

It is important to study changes in Black-White test score gap because the gap substantially accounts for Black-White inequality in socioeconomic status. With respect to racial inequality in educational attainment, Blacks, on average, are less likely to complete college than are Whites. However, when we compare Blacks and Whites with the same test scores, Blacks are slightly more likely than Whites to complete college (Jencks and Phillips, 1998). With respect to racial inequality in earnings, according to Johnson and Neal (1998), the earnings gap between Blacks and Whites for women is entirely explained by the Black-White test score gap. For men, the earnings gap between Blacks and Whites is reduced by half when test scores are equalized. Thus, reduced racial differences in test performance would substantially reduce racial inequality in socioeconomic status.

II. Literature on Long-Term Trends

Virtually all previous studies of changes in the Black-White test score gap have not been based on (a) adults of all ages in the population, (b) identical test instruments over time, or (c) a time span longer than 30 years. Fortunately there is one data source,

the GSS, which is based on a series of nationally representative samples of adults in the U.S. with an identical test instrument that allows the estimation of very long term intercohort trends in racial differences in test scores of a 10-item vocabulary test (WORDSUM).

Using aggregate, annual WORDSUM data from the GSS, 1974-1996, Lynn (1998) regards the GSS vocabulary test as a measure of intelligence and maintains that the Black-White intelligence difference in the U.S. has been persistent over time. However, using data from the GSS, 1974-1994, Huang and Hauser (1998) conclude that Black-White differences in GSS vocabulary test scores have decreased over time. Lynn's (1998) study and Huang and Hauser's (1998) study are not mutually cited and addressed. Three explanations may possibly account for the discrepancy in findings:

First, Lynn conducts a period (survey year) analysis, while Huang and Hauser conduct a cohort analysis.[2] We maintain that cohort analysis is the more direct and appropriate research design; it is essential for an accurate understanding of temporal changes in the Black-White test score gap, if there are any changes. This is not only because verbal ability tends to stabilize after young adulthood (Schaie, 1996: 135), but also because verbal ability, to a large degree, is dependent on experience before adulthood, such as socioeconomic background and

[2] Ryder (1968) defines cohort analysis as "...quantitative description of data occurrences from the time a cohort is exposed to the risk of such occurrences." Period analysis, on the other hand, "involves a quantitative description and analysis of the data for the many cohorts observed during a specified time interval, such as a year, with respect to some variable" (Shryock *et al*. 1976).

schooling. During the twentieth century, not only have average schooling levels increased significantly across cohorts, but also Black-White schooling differentials have decreased across cohorts. A monotonic trend should also appear in a cross-period analysis, yet it requires more cases and more periods to reveal what cohort analysis can show. In the GSS 1974-1998, for instance, there are at most 25 survey years—in several of which no GSS was conducted—but 69 birth cohorts are represented. Thus, the period analysis narrows attention to aggregate trend over a 25-year period. We expect that a period analysis using the GSS data will be biased toward not detecting any trend even when there is one. The cohort replacement process in the period analysis may not have been long enough to reveal intercohort changes that are, by all theoretical accounts, the causal locus of verbal score changes. It requires many more waves of data and a larger sample size to reveal what may be sharp intercohort trends. In short, the period analysis is an inefficient way to detect temporal changes in verbal ability, and it would obscure changes over time in Black-White test score differences. Nevertheless, cohort analysis, as conducted by Huang and Hauser (1998), is based on an assumption that there are no period influences on vocabulary scores.[3] Thus, we think that it is useful to conduct both period and cohort analyses to confirm a convergence over time in the Black-White verbal test score gap if it exists.

Two other possible sources of inconsistent findings are related to statistical strategies. Huang and Hauser (1998) apply

[3] Hauser and Huang (1997: 344-345) investigated the plausibility of this assumption.

sampling weights, while Lynn (1998) does not. Weighting is crucial for descriptive models, such as temporal trends in the gross racial gap in verbal test scores, because unweighted estimates are invalid with mis-specified models (Winship and Radbill, 1994). In addition, Lynn (1998) underestimates the magnitude of the convergence in the Black-White verbal test score gap by failing to correct for a ceiling effect on the GSS vocabulary test scores for Whites in the general population.

Finally, Lynn (1998) adopts a 2-step regression which is unnecessary and builds in an extra layer of possible measurement error. The conventional and well-justified approach to studying group difference and temporal change in the difference is to do standard regression analysis in one single step, as Huang and Hauser (1998) have done.

In sum, Lynn (1998) should have conducted a cohort analysis to address his research question. Lynn (1998) not only has incorrectly chosen to conduct a period analysis, but also has improperly executed it.

Huang and Hauser's (1998) cohort analysis has some limitations which can be improved. First, the most dramatic convergence in verbal ability between Blacks and Whites occurred in the youngest cohort which has a very small sample size for Blacks (N = 28). The use of additional data from the 1996 and the 1998 GSS increases the number of cases in the youngest cohort of Blacks (N = 93) and provides a more reliable estimate of the WORDSUM gap between Blacks and Whites in that youngest cohort. Second, Huang and Hauser (1998) did not include a simple trend model to summarize and test the significance of the magnitude of the narrowing gap. Third, Huang and Hauser (1998) did not examine the extent to which

the regression results are affected by applying sampling weights. Finally, Huang and Hauser (1998) did not conduct a period analysis to confirm a convergence over time in the Black-White verbal test score gap.

Specifically, the objectives of this paper are to: (1) conduct a survey year analysis; (2) conduct a cohort analysis based on a parsimonious regression model, with and without applying sampling weights, and; (3) extend Huang and Hauser's (1998) original intercohort analysis of Black-White differences in WORDSUM for two more rounds of the survey, that is, using data from the GSS from 1974-1998.

III. Data and Methods

Our analysis is based on data from the GSS of the National Opinion Research Center (NORC), which has administered the same 10-item multiple-choice vocabulary test to respondents in the GSS survey years of 1974, 1976, 1978, 1982, 1984, 1987, 1988, 1989, 1990, 1991, 1993, 1994, 1996, and 1998.[4] Because there are repeated 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 data from the GSS, 1974-1998, we can estimate trends in verbal ability among Blacks and Whites who were born from 1909 to 1978.

We urge caution in the use of WORSUM. First, the test is short, so its reliability is low. The internal consistency

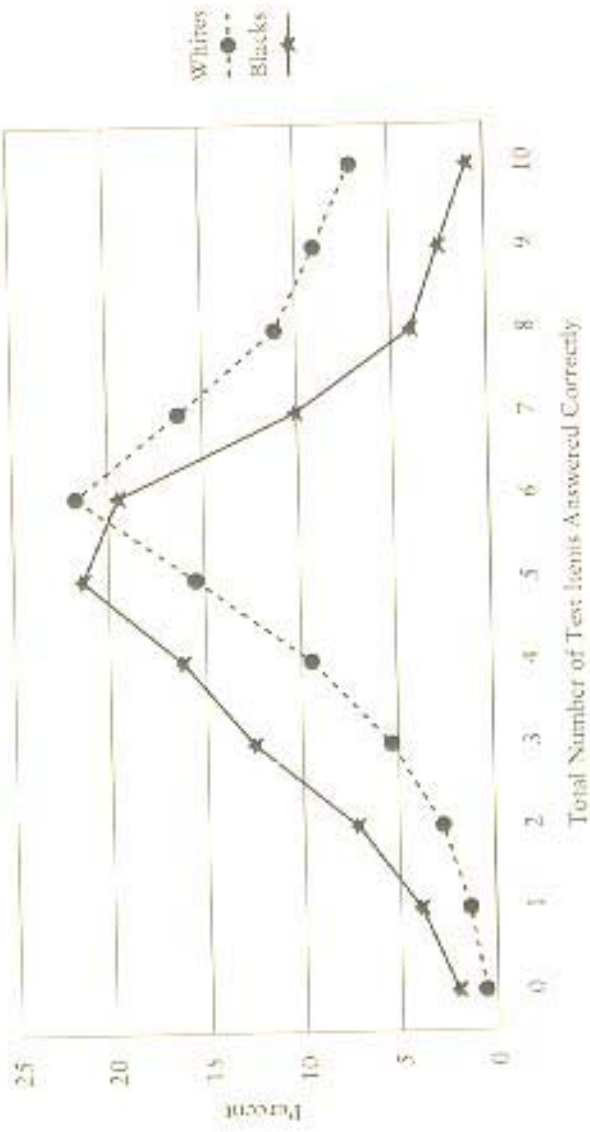
[4] See Huang and Hauser (1998: 304-311) and Hauser and Huang (1997: 340-345) for a description and analyses of the properties of WORDSUM.

reliabilities are about 0.71 among Whites and 0.63 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). The correlations between WORD- SUM and the WAIS-R similarity scores were 0.39 for Whites and 0.41 for Blacks, without correction for unreliability. After correction for internal consistency reliability, the correlations were 0.59 among Whites and 0.61 among Blacks. More specifically, Wolfle (1980: 110) reports that the correlation between the 10-item GSS vocabulary test and the Army General Classification Test (AGCT) aptitude test is 0.71.[5] Second, despite the common use of vocabulary tests in IQ instruments and the high correlations between tests of vocabulary and of general intelligence (Miner, 1957), there is evidence of divergent trends between IQ and verbal ability in the past several decades (Flynn, 1994). Thus, even if some researchers suggest that verbal ability is a valid indicator of general intelligence (Miner, 1957: 28; Jensen, 1980), 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.

The distribution of GSS vocabulary scores which, as displayed in Figure 1, tend to be right-hand censored for Whites. For Blacks, the percentage distribution of vocabulary scores is relatively closer to a normal distribution. Therefore, in our

[5] The AGCT contains four subtests: verbal, arithmetic computation, arithmetic reasoning, and pattern analysis (Zeidner and Drucker, 1983).

Figure 1
Percentage Distribution of the 10-Item Vocabulary Test Scores
Ages 24-89; General Social Survey, 1974-1998



There are 15,636 Whites and 2,692 Blacks in the sample.

cohort analysis and survey year analysis, we adopt a two-sided Tobit specification to correct for the censored distribution (Maddala, 1983). When the censored distribution of WORDSUM is not corrected, the average size of the Black-White gap in WORDSUM is 1.43 items in the 1974-1998 GSS. In addition, the size of the Black-White gap in WORDSUM decreases from 1.57 items in 1974 to 1.12 items in 1998. After correcting for the censored distribution of WORDSUM, the average size of the Black-White gap is 1.49 items from 1974 through 1998. Furthermore, the size of the Black-White gap in WORDSUM decreases from 1.66 items in 1974 to 1.16 items in 1998. Thus, the magnitude of convergence in WORDSUM from 1974 to 1998 between Blacks and Whites is 9.4% larger when censoring of the test-score distribution is taken into account.

IV. The Period Perspective

A. Lynn's Approach

The data for Lynn's (1998) study come from the GSS, 1974-1996. Lynn (1998: 1001) uses "*d* scores" to express Black-White differences in vocabulary scores. The year-specific *d* scores denote the differences between the mean scores of Whites and the mean scores of Blacks, divided by the pooled across-year standard deviation. According to Lynn (1998), the Pearson correlation between GSS survey year and the size of Black-White difference (*d* value) is $r = -0.32$ ($p = 0.28$), which implies a narrowing gap but is not statistically significant. Our replication of Lynn's analysis, based on the unweighted estimates of *d*, yields a higher correlation of $r = -0.41$ ($p = 0.17$),

but the correlation is still not statistically significant.[6] However, when data are weighted, which Lynn (1998) has not done, the Pearson correlation between the GSS survey year and the size of Black-White difference (*d* value) reaches -0.55 ($p = 0.052$), reaching statistical significance at the 0.03 level in a one-tailed test, but not quite at $p = 0.05$ in a two-tailed test (If alternative hypothesis were convergence, one could properly choose an one-tailed test). When data are weighted, and when data from the 1998 GSS are added, a significant correlation of $r = -0.64$ ($p = 0.01$) suggests that the Black-White difference in WORDSUM narrowed over the 24 year period from 1974 to 1998.

Lynn (1998) also regresses *d* values on year of survey, and we have done the same. When weighting is applied, year of survey is found to have marginally significant negative effects (at $p < 0.05$ level) on the size of Black-White difference in WORDSUM. Lynn (1998) does not find a narrowing gap across the 22-year period between 1974 and 1996, because the data are not weighted. But why does the narrowing gap in the size of Black-White difference in WORDSUM become marginally significant when data are weighted? We will address this question later.

B. A Direct Regression Approach to Period Trends

A more rigorous way to test whether there has been a statistically significant narrowing gap between Blacks and Whites in vocabulary scores across survey year is not to use *d*

[6] The detailed results of replicating Lynn's (1998) analysis are presented in Appendix A.

scores and then regress *d* scores on year of survey, but to regress vocabulary scores on race (Blacks = 0, Whites = 1), year of survey, and the interaction of race and year of survey. If the regression coefficient of the interaction term is significantly negative, it means that there has been a statistically significant convergence over time between Blacks and Whites in GSS vocabulary scores.

Table 1 reports the regressions of GSS vocabulary test scores on year of survey, race (Blacks = 0, Whites = 1), and the interaction of year of survey and race. The regression results presented in Table 1 are categorized by whether or not weighting is applied, whether or not the Tobit specification is used to account for the censored distribution of WORDSUM, and whether or not data from the 1998 GSS are incorporated. The interaction effects are uniformly negative, and they are statistically significant even in two-tailed tests at the 0.05 level in weighted analysis of the full data series, with or without correction for censoring.[7] The same conclusion can be reached when year of survey is treated as a series of one-year dummy variables.[8] Contrary to Lynn's conclusion, these findings, too, suggest that the Black-White gap in GSS vocabulary scores narrowed over the survey years from 1974 to 1998.

Nevertheless, results presented in Table 1 suggest, as we expected, that survey year analysis is less efficient and tends to

---

[7] Without applying sampling weights, we treat the weight variable, the product of ADULTS and OVERSAMP, as one of the independent variables. When this is done, we also find the interaction effects are statistically significant even in two-tailed tests at the 0.05 level in weighted analysis of the full data series, with or without correction for truncation.

[8] These regression results are not reported here, but they are available from the authors.

Table 1 Regression of GSS Vocabulary Test Scores on Race, Year of Survey, and the Interaction of Race and Year of Survey: GSS, 1974-1996, and 1974-1998

	Unweighted		Weighted	
	b	*SE*	*b*	*SE*
OLS, GSS, 1974-1996				
Intercept	3.416***	0.565	2.994***	0.586
Whites (Blacks=0)	2.185***	0.602	2.606***	0.620
Year	0.015*	0.007	0.020**	0.007
Whites*Year	-0.0083	0.0070	-0.0132	0.0072
Tobit, GSS, 1974-1996				
Intercept	3.359***	0.602	2.936***	0.624
Whites (Blacks=0)	2.374***	0.642	2.797***	0.661
Year	0.016*	0.007	0.021**	0.007
Whites*Year	-0.0096	0.0074	-0.0147	0.0076
OLS, GSS, 1974-1998				
Intercept	3.095***	0.513	2.715***	0.536
Whites (Blacks=0)	2.419***	0.548	2.846***	0.568
Year	0.019**	0.006	0.023***	0.006
Whites*Year	-0.0111	0.0063	-0.0162*	0.0065
Tobit, GSS, 1974-1998				
Intercept	3.021***	0.547	2.633***	0.571
Whites (Blacks=0)	2.609***	0.585	3.011***	0.606
Year	0.020**	0.006	0.024***	0.007
Whites*Year	-0.0125	0.0067	-0.0173*	0.0069

Note: In order to weight the data, 11 missing cases of ADULTS are deleted. This deletion results in a total of 17,112 cases for the analysis using data from the GSS, 1974-1996. For the analysis using data from the GSS, 1974-1998, this deletion yields a total of 18,337 cases. The weighted and unweighted results are based on the same sample. *$p<0.05$ **$p<0.01$ ***$p<0.001$

be biased toward not detecting any trend even when there is one. The significance of the narrowing Black-White verbal test score gap is marginal, depending on whether or not sampling weights are applied and whether or not data from the 1998 GSS are incorporated.

C. Why Weighting Makes a Difference

Why does the narrowing Black-White gap over time in GSS vocabulary scores become statistically significant when data are weighted? The weight variable applied in this analysis contains two components: the number of adults in the household (ADULTS) and the over-sampling of Blacks in 1982 and 1987 (OVERSAMP). When data are weighted separately by ADULTS and OVERSAMP, as displayed in Table 2, we find that it is the number of adults in the household, not the over-sampling of Blacks in 1982 and 1987, that contributes to the significant Black-White convergence in vocabulary scores. We examine the mean and the standard deviation of WORDSUM by survey year, with and without applying weights.[9] The narrowing Black-White gap over time in WORDSUM is entirely attributable to the upward trend among Blacks. We find that weighting tends to slightly lower the mean vocabulary scores. This is because respondents who live in larger households (in terms of number of adults living in the household) tend to have lower vocabulary scores.[10] That is, when data are weighted so that samples will be

[9] The mean and the standard deviation of WORDSUM by survey year with and without applying weights are presented in Table B1.

[10] The correlation between WORDSUM and ADULTS is -0.057 ($p < 0.0001$) for Whites. For Blacks, it is -0.065 ($p < 0.001$).

Table 2 **Regression of GSS Vocabulary Test Scores on Race, Year of Survey, and the Interaction of Race and Year of Survey: GSS, 1974-1998**

Model	*b*	*SE*
Tobit, GSS, 1974-1998, unweighted, N=18,337		
Intercept	3.021***	0.547
Whites (Blacks=0)	2.609***	0.585
Year	0.020**	0.006
Whites*Year	-0.0125	0.0067
*Tobit, GSS, 1974-1998, weighted by OVERSAMP*ADULTS, N=18,337*		
Intercept	2.633***	0.571
Whites (Blacks=0)	3.011***	0.606
Year	0.024***	0.007
Whites*Year	-0.0173*	0.0069
Tobit, GSS, 1974-1998, weighted by ADULTS, N=18,337		
Intercept	2.643***	0.549
Whites (Blacks=0)	3.015***	0.585
Year	0.024***	0.006
Whites*Year	-0.0171*	0.0067
Tobit, GSS, 1974-1998, weighted by OVERSAMP, N=18,337		
Intercept	2.995***	0.571
Whites (Blacks=0)	2.619***	0.607
Year	0.020***	0.006
Whites*Year	-0.0127	0.0069
*Tobit, GSS, 1974-1998, N=18,337, weighted by OVERSAMP*ADULTS for Whites in 1974-1998 and Blacks in years 1984-1998, weighted by OVERSAMP for Blacks in 1982, no weighting applied for Blacks in years 1974-1978*		
Intercept	2.811***	0.704
Whites (Blacks=0)	2.832***	0.732
Year	0.022**	0.008
Whites*Year	-0.0154	0.0082

*$p<0.05$ **$p<0.01$ ***$p<0.001$

representative of the U.S. population on the individual level, the share of lower scorers increases. Weighting tends to reduce the mean vocabulary scores of Blacks more in earlier survey years, such as in 1974, 1976, 1978, and 1982, than it does in later years. That is, when data are weighted, the upward WORDSUM trend of Blacks becomes steeper, because the upward trend starts from a lower level. Therefore, when data are weighted, the narrowing Black-White gap in WORDSUM becomes more statistically significant. To test this hypothesis, we examine whether or not the significant effects of the interaction of race and survey year would become insignificant, when data are weighted by OVERSAMP*ADULTS, but no weighting is applied to Blacks in 1974-1978, and only OVERSAMP is used to weight the Black sample in 1982 (because Blacks were oversampled in that year). The regression results reported in the last panel of Table 2 support the hypothesis because the interaction effect is reduced from 0.0173 to 0.0154, and the latter coefficient becomes insignificant.

V. The Cohort Perspective

In order to conduct the intercohort analysis, several case selections are made.[11] The total number of unweighted cases is 12,755 for Whites and 2,308 for Blacks after selection on valid cases. For example, with respect to the White sample, the cumulative sample size in the GSS, 1974-1998, is 31,845, but the 10-item vocabulary test was not administered every year (years administered include: 1974, 1976, 1978, 1982, 1984,

[11] Table B2 shows the details of case selections.

1987-1991, 1993, 1994, 1996, and 1998), thus reducing the potential sample size in this analysis to 21,206. Also, due to a switch of rotation design from across surveys to within surveys, one-third of the respondents are "Not Applicable" for all 10 word items after 1988; these "Not Applicable" cases further reduce the potential sample size in this analysis to 16,379. To construct a variable, CWORDSUM (as we name it to indicate "Corrected WORDSUM") which represents the total number of correct word items, we take those respondents who did not answer all test items, but did answer some test items, to have answered the missing items incorrectly. As it appears in the GSS 1974-1998 codebook, the WORDSUM variable, which purports to measure the total number of correct word items, was constructed in this way; however, there are some coding errors in the construction of the WORDSUM variable.[12] Selection on valid cases of the variable, CWORDSUM, excludes 703 respondents who had "No Answers" on all test items. We include only persons who are at least 20 years old, because many younger people are still in school. Based on Cattell's (Cattell, 1971a, 1971b) hypothesis that "crystallized intelligence" starts to decline at about 60 to 70 years of age and that vocabulary

[12] It is expected that if all word items are "No Answers", they should be in the category of WORDSUM = 99. However, 31 cases out of 907 "No Answers" cases in all word items in the entire sample of GSS, 1974-1994, are in WORDSUM = 0. It is also expected that respondents in the category of WORDSUM = 99 should not have answered any item correctly or incorrectly, but should have no answer. However, 24 respondents in WORDSUM = 99 answered WORDH incorrectly, and 22 respondents in WORDSUM = 99 answered WORDH correctly in the entire sample of GSS, 1974-1998. To obtain the correct measure of total number of correct word items, we constructed "CWORDSUM", referring to corrected WORDSUM.

tests, when taken at age 65 or older, may not reflect past verbal ability, we limit our analysis to persons under age 66. Therefore, selection of the age range 20-65 excludes 2,868 cases. Selection on valid cases of the "number of siblings" variable excludes 21 cases; selection on valid cases of the variable "Not living with both own parents at age 16" excludes 4 cases; selection on valid cases of the "number of adults in the household" variable excludes 8 cases; selection on valid respondent's years of education excludes 20 cases. These selections reduce the sample size to 12,755 cases for Whites. Going through the same case selections, the sample size is reduced to 2,308 cases for Blacks. The total number of cases is 15,063.

A. Reconsidering Intercohort Trends

To examine the significance of the narrowing Black-White verbal score gap over time, we estimate three models of intercohort trend in verbal ability, based on a two-sided Tobit specification, which takes the censored distribution of the GSS vocabulary test scores into account. In Model 1, we regress the total number of correct answers to the 10 CWORDSUM items on race, age, birth year, and interaction between race and birth year:

$$E\left[y\right] = \alpha + \sum_{1}^{I} \beta_i X_i + \gamma_1 W_1 + \delta_1 Z_1 + \lambda_1 Z_1 W_1, \qquad (1)$$

where y is the number of correct CWORDSUM items, α is the intercept, the X_i are dummy variables for age groups, the β_i are age effects, the W_1 is a continuous variable for birth year which ranges from 0 (when respondents' birth year is in 1909) to 69 (when respondents' birth year is in 1978), the γ_1 are

effects of birth year, Z_1 is a dummy variable for race, the δ_1 are effects of race, and the λ_1 are effects of the race by birth year interaction.

In Model 1, we estimate intercohort trends in verbal ability among Blacks and Whites, net of an age effect. The estimated coefficients and their standard errors are reported in Table 3. A statistically significant negative coefficient in the interaction of birth year and race indicates a convergence in the Black-White test score gap over time. In Model 2, we add eight social background measures to Model 1 of Equation 1. These measures include 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, Southern residence (at age 16), and three dummy variables that flag missing values on father's education, mother's education, and father's occupation. In Model 3, we add respondent's years of schooling to Model 2. Years of education are coded into a series of 21 dummy variables with 12 years of education as the reference group.

The results of the cohort analysis, as presented in Model 1 of Table 3, suggest that the narrowing Black-White verbal test score gap over time is substantial and statistically significant, with or without applying sampling weights ($p < 0.001$). Even after controlling for respondent's social background and years of schooling completed, as presented in the Model 3 of Table 3, the magnitude of the Black-White convergence drops by half, but it remains statistically significant at the level, $p < 0.01$, with or without applying sampling weights.

In the survey year analysis, findings vary, depending on

Table 3 Three Regression Models of Vocabulary Test Scores: General Social Survey, 1974-1998

	Model 1		Model 2		Model 3	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*
Unweighted						
Intercept	4.02***	0.18	4.12***	0.18	5.40***	0.17
Whites (Blacks=0)	2.12***	0.15	1.43***	0.14	1.20***	0.13
Age						
20-29						
30-39	0.54***	0.05	0.50***	0.05	0.23***	0.04
40-49	0.71***	0.07	0.65***	0.06	0.28***	0.05
50-59	0.44***	0.09	0.37***	0.08	0.02	0.07
60-65	0.23*	0.11	0.15	0.10	-0.12	0.09
Birth year	0.012**	0.004	-0.013***	0.004	-0.026***	0.003
Whites*Birth Year	-0.017***	0.004	-0.017***	0.003	-0.009**	0.003
Weighted						
Intercept	3.90***	0.19	4.09***	0.19	5.37***	0.17
Whites (Blacks=0)	2.24***	0.16	1.47***	0.15	1.23***	0.14
Age						
20-29						
30-39	0.54***	0.05	0.51***	0.05	0.24***	0.04
40-49	0.66***	0.06	0.64***	0.06	0.27***	0.05
50-59	0.41***	0.09	0.37***	0.08	0.01	0.07
60-65	0.23*	0.11	0.18	0.10	-0.12	0.09
Birth year	0.015***	0.004	-0.011**	0.004	-0.025***	0.003
Whites*Birth Year	-0.020***	0.004	-0.018***	0.004	-0.010**	0.003

Note: Age effects are expressed relative to ages 20-29. In Model 1, we estimate intercohort trends in verbal ability among Blacks and Whites, net of an age effect. In Model 2, we add eight social background measures to Model 1. In Model 3, we add respondent's years of schooling to Model 2. Years of education are coded into a series of 21 dummy variables with 12 years of education as the reference group. Regression coefficients and their standard errors for eight social background variables and 21 education dummy variables are available from the authors. *$p<0.05$ **$p<0.01$ ***$p<0.001$

whether or not sampling weights are applied and whether or not data from the 1998 GSS are incorporated. In the cohort analysis, the convergence in the Black-White test score gap over time is very significant and substantial, whether or not sampling weights are applied.

To display the intercohort trends in GSS vocabulary test scores of Blacks and Whites, we adopt Huang and Hauser's (1998) model specification and estimate three models of intercohort trend in verbal ability, based on a two-sided Tobit specification which takes the censored distribution of the GSS vocabulary test scores into account. In the baseline model, we regress the total number of correct answers to the 10 CWORDSUM items on sex, race, age, birth cohort, and interaction between race and sex and between race and birth cohort:[13]

$$E[y] = \alpha + \sum_1^I \beta_i X_i + \sum_1^J \gamma_j W_j + \delta_1 Z_1 + \delta_2 Z_2 + \delta_3 (Z_1 Z_2) + \sum_1^J \lambda_j Z_1 W_j, \quad (2)$$

where y is the number of correct CWORDSUM 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 δs are effects of sex and race, and the λ_j are effects of race by cohort interactions. In the social

[13] The number of cases for Whites born in 1909-1919, 1920-1929, 1930-1939, 1940-1949, 1950-1959, 1960-1969, and 1970-1978 are 543, 1304, 1876, 2,974, 3,426, 2,107, and 525, respectively. The counts for Blacks were 60, 186, 328, 501, 697, 443, 93, respectively. The small number of cases in the oldest and the youngest cohorts of Blacks suggests that findings about them should be taken with caution.

background model, we add eight social background measures (as mentioned above) to the baseline model of Equation 2. In the background & schooling model, we add respondent's years of schooling to the social background model.[14] Again, years of schooling are coded into a series of 21 dummy variables with 12 years of schooling as the reference group.

Figure 2 shows the intercohort trends in GSS vocabulary test scores of Blacks and Whites, as estimated in the baseline model. The Black-White gap in verbal ability has been narrowing.[15] Relative to the earliest birth cohort, the Black-White differences in WORDSUM are 84%, 65%, 62%, 64%, 49%, and 24% as large in successive cohorts.[16] The Black-White convergence is monotonic; it shows no sign of ending. Furthermore, about three-fourths of the convergence is attributable to an upward intercohort trend among Blacks, from 4.1 in the earliest cohort to 5.4 in the most recent cohort. About one-fourth of the convergence is due to a downward intercohort trend among Whites, from 6.4 in the earliest cohort to 6.0 in the most recent cohort. The significance of the Black-White convergence, the fact that the convergence is continuous, and

[14] The detailed regression results of the baseline, social background, and background & schooling models are presented in Table B3.

[15] In the baseline model, the likelihood-ratio test statistics for intercohort differences in test scores and for the Race × Cohort interactions are both highly significant.

[16] Because of the small number of cases in both the oldest and the youngest cohorts of Blacks, one should take these observations with caution. The Black-White convergence, however, is still highly significant, when the two extreme cohorts (1909-1919 and 1970-1978), either Blacks or Whites, are excluded entirely from the analysis. In the 1960-1969 birth cohort, the Black-White difference is 58.7% as large as in the 1920-1929 birth cohort.

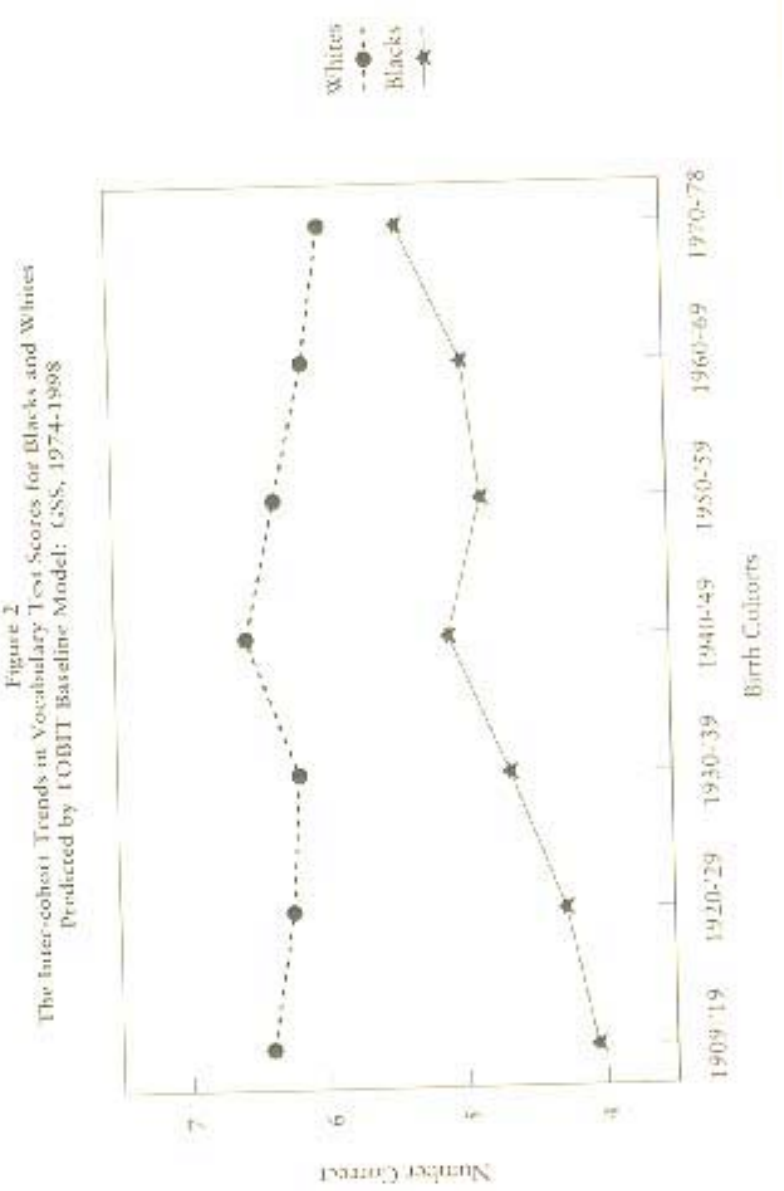


Figure 2
The Inter-cohort Trends in Vocabulary Test Scores for Blacks and Whites
Predicted by 1OBIT Baseline Model: GSS, 1974-1998

the fact that most of the convergence is attributable to improvement in Blacks' test scores, are all very encouraging.

B. Explaining the Convergence

To understand why Black-White differences in test scores change over time, it is helpful to examine the extent to which the Black-White gap in GSS vocabulary test scores is attributable to Black-White differences in social background and amount of schooling, and the extent to which improvements in social background and schooling across time contribute to the Black-White convergence. Figure 3 displays intercohort trends in GSS vocabulary test scores for Blacks and Whites, as estimated in the social background model. After controlling for social background, the upward intercohort trend among Blacks becomes flat. This implies that the observed upward intercohort trend for Blacks is related to an improvement in social background. For Whites, however, improvement in social background across time does not raise test scores correspondingly. A favorable social background for Whites in the later cohorts does not raise their scores to the degree that scores of Whites in the earlier cohorts would be raised if they had the same favorable social background as that of the later cohorts. Therefore, the intercohort trend for Whites is downward after controlling for social background. One alternative explanation is that indicator variables of social background, such as father's years of education and mother's years of education, have increased quantitatively over time, but the selectivity of education has decreased (Wilson and Gove, 1999a, 1999b; Glenn, 1999;

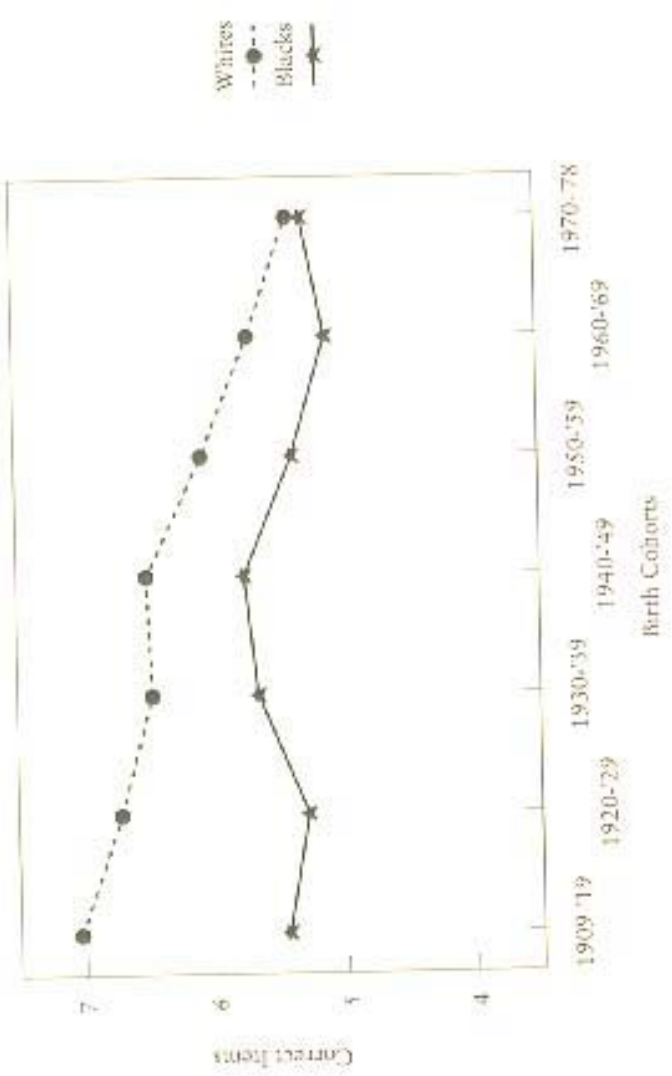


Figure 4
The Inter-cohort Trends in Vocabulary Test Scores for Blacks and Whites
Predicted by TOBIT Social Background Model: GSS, 1974-1998

Alwin and McCammon, 1999).[17] It could also be that the quality of education has decreased (Hayes, Wolfer, and Wolfe, 1996). Most importantly, when social background is controlled, Black-White differences are significantly reduced; the Black-White convergence in verbal ability also becomes less rapid, but it remains statistically significant.[18]

Figure 4 shows the intercohort trends in test scores when respondents' years of schooling as well as social background are added to the baseline model. After controlling for both social background and schooling, the Black-White differences in the earliest two birth cohorts are substantially reduced; the differential is more than one correct answer on the WORDSUM test (slightly more than half a standard deviation), and in all but the youngest cohort, the differential is slightly less than one correct answer. The intercohort trends of Blacks and Whites are both downward. The magnitude of the Black-White convergence across birth cohorts, though reduced, continues to be statistically significant, apparently due to the relatively high performance of the youngest cohort of Blacks.[19]

---

[17] An education selection effect exists because, given the same education level, recent-borns tend to be less selective with respect to scholastic ability than earlier borns, as the average years of education completed has increased markedly over time.

[18] In the social background model, the likelihood-ratio test statistics for intercohort differences in test scores and for the Race X Cohort interactions are both highly significant.

[19] In the background & schooling model, the likelihood-ratio test statistics for intercohort differences in test scores are highly significant. For the Race X Cohort interactions, the likelihood-ratio test statistics remain to be significant at $p < 0.005$ level. However, when the most recent cohort, 1970-1978, is excluded from the analysis, the Race X Cohort interactions become insignificant.

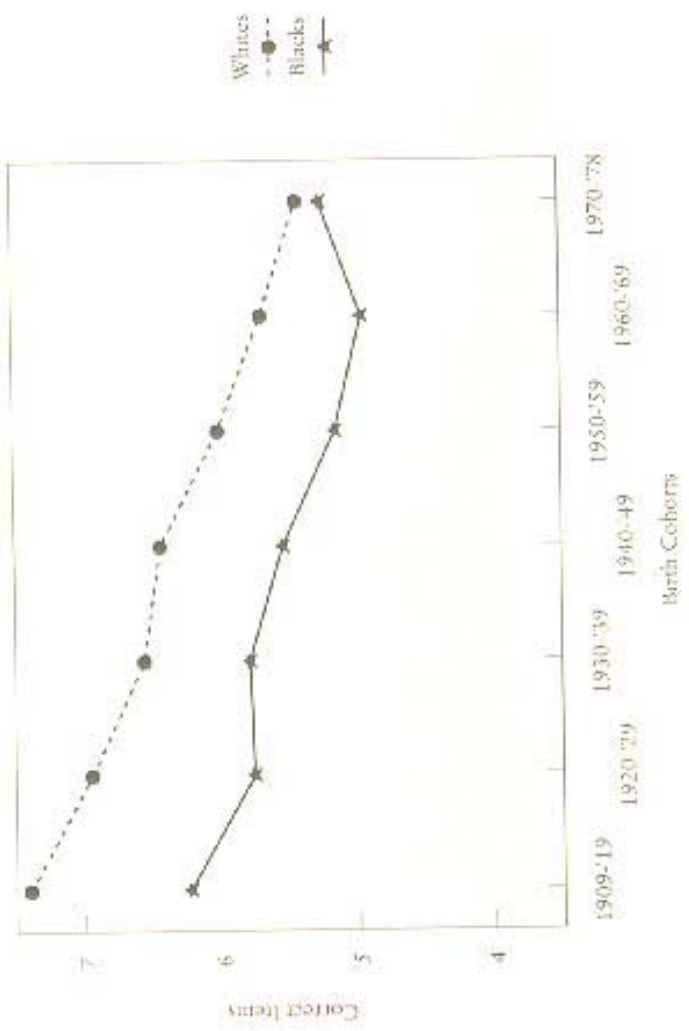


Figure 4
The Inter-cohort Trends in Vocabulary Test Scores for Blacks and Whites
Predicted by TOBIT Background & Schooling Models: GSS, 1974-1998

On average, Black-White differences in social background and schooling account for 45% of Black-White differences in GSS vocabulary test scores. That is, changes in social background and schooling contribute to the convergence in test scores. For example, in the baseline model, the change in Black-White difference between the oldest and the youngest cohorts was 1.8 points; in the social background model, it was 1.5 points; and in the background & schooling model, it was 1 point. Nevertheless, there are additional signs of convergence in the youngest cohort due to the relatively high performance of the youngest cohort of Blacks. Except for the most recent birth cohort, 1970-1978, cohort changes in social background and schooling account for the narrowing over time of the gap between Blacks and Whites in GSS vocabulary test scores. The civil rights movement and War on Poverty initiated in the late 1960s and early 1970s might have contributed to the relatively high performance of the youngest cohort of Blacks, those born between 1970 and 1978 (Grissmer, Flanagan, and Williamson, 1998).

VI. Conclusion

Using data from the GSS, Lynn (1998) finds a persistent Black-White test score gap over time, but Huang and Hauser (1998) report significant Black-White test score convergence. While Lynn (1998) analyzes changes in Black-White test score differences across survey year, Huang and Hauser (1998) evaluate changes in Black-White test score differences across birth cohorts. Since the real action for changes over time in Black-White test scores differences is in cohort replacement,

cohort analysis is a more sensible strategy. Cohort analysis not only captures the significance of changes over time in Black-White test score differences, but also allows for a Black-White comparison based on a longer time span.

Our analysis of intercohort changes in Black-White vocabulary test score differences suggests that the Black-White test score gap has significantly narrowed over time. The narrowing, in large part, is due to an upward trend among Blacks, whose social background and schooling have improved. The survey year analyses we conducted also show that, (1) when data from the 1998 GSS are incorporated and (2) when data are weighted appropriately, the Black-White test score gap narrowed from 1974 to 1998. Therefore, we cannot agree with Lynn's conclusion that the GSS data provide "no conclusive evidence that the black-white difference in intelligence has been narrowing over time." No one could claim that the present analysis offers "conclusive" evidence about broader measures of "intelligence" than the vocabulary task measured in the GSS. However, we find that Lynn's analysis of the GSS data is faulty. Proper analysis of the GSS data, either of aggregate trend across survey years or across birth cohorts, provides consistent and significant evidence that vocabulary differences between Black and White Americans narrowed through most of the 20th century.

MIN-HSIUNG HUANG is an assistant research fellow in sociology at the Institute of European and American Studies, Academia Sinica. His areas of interest include social stratification, sociology of education, and survey research methodology. He is currently working on a project to examine the intensity of cognitive partitioning by occupation. He is also writing a proposal to study the learning effects of public screening and evaluative examinations.

ROBERT M. HAUSER is Vilas Research Professor of Sociology at the University of Wisconsin-Madison, where he has directed the Center for Demography and Ecology and the Institute for Research on Poverty. He currently directs the Center for Demography of Health and Aging, which is supported by the National Institute on Aging. He is a member of the National Academy of Sciences and is a fellow of the National Academy of Education, the American Association for the Advancement of Science, the American Statistical Association, the Center for Advanced Study in the Behavioral Sciences, and the American Academy of Arts and Sciences. He has served on the National Research Council's Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, and Board on Testing and Assessment. He chaired the National Research Council's Committee on the Appropriate Test Use. His current research interests include trends in educational progression and social mobility in the United States among racial and ethnic groups, the uses of educational assessment as a policy tool, the effects of families on social and economic inequality, and changes in socioeconomic standing, health, and well-being across the life course.

Appendix A: Replicating Lynn's Analysis

The data for Lynn's (1998) study come from the GSS, 1974-1996. Lynn reports there are 14,657 Whites and 2,510 Blacks. While we have reproduced his count for Blacks, we cannot reproduce his exact count for Whites. We contacted Richard Lynn by letter to learn how he had selected GSS cases, and in March 1999, he wrote to say that he had discarded his work with the WORDSUM data and could not respond to our queries.

Starting with the full file and excluding individuals to whom WORDSUM was not administered, we count 15,281 Whites and 2,807 Blacks, for a total of 18,088. There are several ways of proceeding from that figure, none of which gives us Lynn's numbers for both Whites and Blacks. First, we deleted those cases in which WORDSUM = 99. This gave us 14,613 Whites and exactly 2,510 Blacks. This method gives us Lynn's count of Blacks, but leaves out 44 Whites. Our total adds up to 17,123, while Lynn computes a total of 17,167. Second, we deleted those cases in which the responses are "9" (no answer) for all 10 word items, recoding 9 as 0 (answered incorrectly) for each word item, and reconstructing WORDSUM by adding up the correct item responses. This gave us 14,633 Whites and 2,504 Blacks, for a total of 17,137. Finally, by following the second procedure, we found that several of the deleted cases have WORDSUM = 0, rather than 99. If we add these cases back in, we get 14,649 Whites and 2,516 Blacks, for a total of 17,165 cases. None of these three case selection procedures yields the exact count for Whites that is recorded in Lynn's sample. To repeat Lynn's analysis, we chose to delete those

cases in which WORDSUM = 99. This yields 14,613 Whites and exactly 2,510 Blacks.

Lynn (1998: 1001) uses "*d* scores" to express Black-White differences in vocabulary scores. The year-specific *d* scores denote the differences between the mean scores of Whites and the mean scores of Blacks, divided by the "pooled standard deviation." Lynn (1998) does not mention whether or not he uses the pooled across-year standard deviation or the pooled within-year standard deviation. Therefore, Table A1 displays five versions of *d* value by survey year. The *d* values in the first column are as reported by Lynn (1998). Unweighted *d* values based on the pooled across-year standard deviation are presented in column 2, while unweighted *d* values based on the pooled within-year standard deviation are presented in column 4. In order for the GSS samples to be representative of the U.S. population on the individual level, and to take into account the over-sampling of blacks in the 1982 and 1987 surveys, it is necessary to weight the data by the product of two variables in the public GSS data file, ADULTS (the number of adults in a household) and OVERSAMP (Davis and Smith, 1992). The *d* values based on weighted means and weighted pooled across-year standard deviations are presented in column 3, while column 5 presents the *d* values based on weighted means and weighted pooled within-year standard deviations. The *d* values in column 2 are closest to Lynn's *d* values; the figures agree to two decimal places in six years (note that none of our four estimates is at all close to Lynn's $d = 0.56$ in 1993). This suggests that Lynn may have incorrectly used the unweighted data to estimate pooled across-year standard deviations.

According to Lynn (1998), the Pearson correlation

between GSS survey year and the size of Black-White difference (*d* value) is $r = -0.32$ ($p = 0.28$), which implies a narrowing gap but is not statistically significant. Note that Lynn's *p*-values are perhaps excessively conservative, based on a two-tailed test. If the alternative hypothesis were convergence, one could properly choose a one-tailed test, yielding a *p*-value half as large as reported by Lynn. Our replication of Lynn's analysis, based on the unweighted estimates of *d*, yields a higher correlation of $r = -0.41$ ($p = 0.17$), but the correlation is still not statistically significant. When data are weighted, however, the Pearson correlation between the GSS survey year and the size of Black-White difference (*d* value) reaches -0.55 ($p = 0.052$), reaching statistical significance at the 0.03 level in a one-tailed test, but not quite at $p = 0.05$ in a two-tailed test. When data are weighted and when data from the 1998 GSS are added, as displayed in Table A2, a significant correlation of $r = -0.64$ ($p = 0.01$) suggests that the Black-White difference in WORDSUM narrowed over the 24-year period from 1974 to 1998. On the other hand, when *d* values are based on the pooled within-year standard deviation, with or without weighting, the magnitude of the narrowing Black-White gap in WORDSUM becomes insignificant. This is because the pooled within-year standard deviation of WORDSUM decreases over time. It decreases from 2.2 in 1974 to 2.0 in 1998, as persons who scored extremely high or extremely low became a proportionally smaller share of the sample. Consistent with Lynn's analysis, we think that the *d* values based on the pooled across-year standard deviation should be used.

Lynn (1998: 1001) also regresses *d* values on year of survey, and we have done the same. The regression results for

the series in Table A1 are presented in Table A3. When weighting is applied and when *d* values are based on the pooled across-year standard deviation, year of survey is found to have significant negative effects on the size of Black-White difference in WORDSUM. Lynn (1998) does not find a narrowing gap across the 22-year period between 1974 and 1996 because the data are not weighted.

Table A1
Differences between Blacks and Whites on Vocabulary Scores Expressed as *d*s, General Social Survey, 1974-1996

		Using Pooled Across-Year SD		Using Pooled Within-Year	
Year	Lynn's *d*	*d* (unweighted)	*d* (weighted)	*d* (unweighted)	*d* (weighted)
1974	0.72	0.72	0.73	0.70	0.71
1976	0.79	0.79	0.84	0.77	0.81
1978	0.70	0.70	0.74	0.69	0.71
1982	0.66	0.66	0.69	0.65	0.69
1984	0.67	0.66	0.67	0.65	0.66
1987	0.73	0.72	0.73	0.70	0.71
1988	0.58	0.57	0.58	0.57	0.58
1989	0.65	0.65	0.67	0.64	0.65
1990	0.46	0.46	0.47	0.46	0.48
1991	0.70	0.66	0.67	0.69	0.70
1993	0.56	0.70	0.71	0.74	0.74
1994	0.83	0.75	0.70	0.80	0.74
1996	0.64	0.60	0.61	0.62	0.63
Year and *d* Correlation	-0.32 (p=0.28)	-0.41 (p=0.17)	-0.55 (p=0.05)	-0.18 (p=0.55)	-0.36 (p=0.23)

Note: The year-specific *d* scores denote the differences between the mean vocabulary scores of Whites and the mean vocabulary scores of Blacks, divided by either the pooled across-year standard deviation or the pooled within-year standard deviation of vocabulary scores.

Table A2

Differences between Blacks and Whites on Vocabulary Scores Expressed as *d*s, General Social Survey, 1974-1998

		Using Pooled Across-Year SD		Using Pooled Within-Year SD	
Year	Lynn's *d*	*d* (unweighted)	*d* (weighted)	*d* (unweighted)	*d* (weighted)
1974	0.72	0.72	0.73	0.70	0.71
1976	0.79	0.79	0.84	0.77	0.81
1978	0.70	0.71	0.74	0.69	0.71
1982	0.66	0.66	0.69	0.65	0.69
1984	0.67	0.67	0.67	0.65	0.66
1987	0.73	0.72	0.73	0.70	0.71
1988	0.58	0.58	0.58	0.57	0.58
1989	0.65	0.65	0.67	0.64	0.65
1990	0.46	0.46	0.47	0.46	0.48
1991	0.70	0.66	0.68	0.69	0.70
1993	0.56	0.71	0.71	0.74	0.74
1994	0.83	0.75	0.70	0.80	0.74
1996	0.64	0.60	0.61	0.62	0.63
1998	—	0.56	0.52	0.59	0.55
Year and *d* Correlation	-0.32 (p=0.28)	-0.49 (p=0.08)	-0.64 (p=0.01)	-0.26 (p=0.37)	-0.47 (p=0.09)

Note: The year-specific *d* scores denote the differences between the mean vocabulary scores of Whites and the mean vocabulary scores of Blacks, divided by either the pooled across-year standard deviation or the pooled within-year standard deviation of vocabulary scores.

Table A3
Regression of *d* Values on Year of Survey, GSS, 1974-1996

Dependent Variable	*b*	*SE*	R^2
GSS, 1974-1996			
Lynn's d	-0.0045	0.0039	0.10
d (unweighted), across-year SD	-0.0049	0.0033	0.16
d (weighted), across-year SD	-0.0069*	0.0031	0.30
d (unweighted), within-year SD	-0.0023	0.0037	0.03
d (weighted), within-year SD	-0.0043	0.0033	0.13

Note: The year-specific *d* scores denote the differences between the mean vocabulary scores of Whites and the mean vocabulary scores of Blacks, divided by either the pooled across-year standard deviation or the pooled within-year standard deviation of vocabulary scores. *$p<0.05$

Appendix B: Additional Tables

Table B1

Mean and Standard Deviation of WORDSUM by Year of Survey With and Without Applying Weighting, General Social Survey, 1974-1998

	N		Unweighted Mean		Weighted Mean		Unweighted SD		Weighted SD	
Year	Blacks	Whites	Blacks	Whites	Blacks	Whites	Blacks	Whites	Blacks	Whites
1974	166	1,277	4.66	6.21	4.59	6.17	1.87	2.19	1.80	2.18
1976	116	1,313	4.47	6.19	4.34	6.14	2.09	2.19	2.08	2.16
1978	147	1,324	4.59	6.12	4.56	6.10	1.92	2.21	1.94	2.19
1982	451	1,252	4.71	6.14	4.64	6.12	2.17	2.10	2.16	2.09
1984	155	1,199	4.77	6.20	4.78	6.22	2.00	2.18	1.99	2.16
1987	465	1,166	4.64	6.19	4.60	6.16	2.01	2.16	1.90	2.18
1988	113	779	4.71	5.95	4.68	5.92	1.92	2.18	1.83	2.12
1989	94	840	4.72	6.12	4.62	6.05	2.06	2.15	2.06	2.18
1990	95	732	5.34	6.33	5.30	6.31	2.28	2.11	2.19	2.09
1991	118	817	4.89	6.31	4.83	6.28	1.76	2.06	1.82	2.04
1993	106	857	4.74	6.26	4.73	6.24	1.93	2.02	1.90	2.00
1994	231	1,539	4.78	6.41	4.87	6.37	1.76	1.99	1.78	1.98
1996	253	1,518	4.98	6.27	4.93	6.24	1.95	2.04	1.96	2.03
1998	182	1,043	5.18	6.38	5.19	6.30	1.84	2.02	1.86	2.03

Table B2

Case Selections for Intercohort Analysis of Changes in Vocabulary Test Score Differences between Blacks and Whites: GSS, 1974-1998

Case selection	Number of Cases	
	Whites	Blacks
The cumulative sample size in the GSS, 1974-1998.	31,845	5,183
Vocabulary test was not administered in 1972, 1973, 1975, 1977, 1980, 1983, 1985, and 1986.	21,206	3,759
"Non Applicable" cases due to a switch of rotation design from across surveys to within surveys since 1988.	16,379	3,009
Excluding "No Answer" to all test items cases.	15,676	2,686
Age 20-65.	12,808	2,331
The valid cases of the "number of siblings" variable.	12,787	2,317
The valid cases of the variable "Not living with both own parents at age 16."	12,783	2,314
The valid cases of the "number of adults in the household variable."	12,775	2,312
The valid cases of the "respondent's years of education completed."	12,755	2,308

Note: Counts are the cumulative numbers of cases remaining after application of the stated criterion.

Table B3
Baseline, Social Background, and Background & Schooling Models of Vocabulary Test Scores: General Social Survey, 1974-1998

	Baseline		Social Background		Background & Schooling	
Variable	*b*	*SE*	*b*	*SE*	*b*	*SE*
Intercept	3.69***	0.31	3.65***	0.29	5.22***	0.26
Men (Women = 0)	0.07	0.11	0.02	0.10	0.06	0.09
Whites (Blacks=0)	2.48***	0.31	1.69***	0.28	1.35***	0.25
Age						
20-29	—	—	—	—		
30-39	0.44***	0.05	0.45***	0.05	0.24***	0.05
40-49	0.56***	0.07	0.58***	0.06	0.30***	0.06
50-59	0.44***	0.08	0.46***	0.08	0.11	0.07
60-65	0.33**	0.11	0.35***	0.10	-0.03	0.09
Birth cohort						
1909-1919	—	—	—	—	—	—
1920-1929	0.22	0.35	-0.16	0.31	-0.47	0.28
1930-1939	0.60	0.33	0.21	0.29	-0.44	0.27
1940-1949	1.03**	0.32	0.31	0.29	-0.69**	0.26
1950-1959	0.77*	0.32	-0.07	0.29	-1.08***	0.26
1960-1969	0.89**	0.33	-0.34	0.29	-1.28***	0.27
1970-1978	1.35***	0.38	-0.16	0.34	-0.98**	0.31
Women*Whites	-0.29*	0.11	-0.24*	0.10	-0.39***	0.09
Cohort*Whites						
1909-1919	—	—	—	—	—	—
1920-1929	-0.38	0.36	-0.16	0.32	0.01	0.29
1930-1939	-0.82*	0.34	-0.77*	0.30	-0.41	0.28
1940-1949	-0.89**	0.33	-0.84**	0.29	-0.28	0.27
1950-1959	-0.85**	0.32	-0.89**	0.29	-0.32	0.26
1960-1969	-1.19***	0.33	-0.98**	0.30	-0.44	0.27
1970-1978	-1.79***	0.39	-1.47***	0.35	-1.00**	0.32
Father's education			0.06***	0.01	0.02***	0.01
Mother's education			0.10***	0.01	0.04***	0.01
Father's occupational status			0.01***	0.00	0.01***	0.00
Number of siblings			-0.08***	0.01	-0.04***	0.01

(to be continued)

Table B3 (Continued)

Not living with both parents at age 16	0.02	0.05	0.10*	0.05
Lived in foreign country at age 16	0.13	0.33	-0.17	0.30
Farm background	-0.45***	0.04	-0.32***	0.04
Lived in the South at age 16	-0.46***	0.04	-0.35***	0.03
Missing father's education	-0.45***	0.06	-0.09	0.05
Missing mother's education	-0.67***	0.06	-0.27***	0.05
Missing father's occupational status	0.19**	0.07	0.03	0.06
No schooling			-0.12	0.68
Years of school				
1			-3.17**	0.98
2			-1.42*	0.66
3			-2.58***	0.35
4			-2.61***	0.32
5			-2.37***	0.26
6			-1.51***	0.20
7			-1.57***	0.14
8			-1.28***	0.09
9			-1.09***	0.09
10			-0.73***	0.07
11			-0.71***	0.07
12			—	—
13			0.38***	0.06
14			0.72***	0.05
15			1.03***	0.07
16			1.64***	0.05
17			1.84***	0.09
18			2.03***	0.09
19			2.44***	0.14
20+			2.42***	0.12

Note: Entries for reference groups in sets of more than two dummy variables are marked by a dash. Thus, main effects of race pertain to the Black-White difference in the 1909-1919 cohort; main effects of cohorts pertain to contrasts among Blacks, relative to the 1909-1919 cohort; and Cohort×Whites interactions pertain to cohort-specific Black-White differences. Age effects are expressed relative to ages 20-29, and education effects are expressed relative to the completion of 12 years of school.
*$p<0.05$ **$p<0.01$ ***$p<0.001$

References

Alwin, Duane F. and Ryan J. McCammon. 1999. "Aging versus Cohort Interpretations of Intercohort Differences in GSS Vocabulary Scores." *American Sociological Review*, Vol. 64, pp. 272-286.

Campbell, Jay R., Catherine M. Hombo, and John Mazzeo. 2000. *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*. U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics. NCES 2000-469, Washington, DC: 2000.

Cattell, Raymond B. 1971a. "The Structure of Intelligence in Relation to the Nature-Nurture Controversy." Pp. 3-30 in *Intelligence: Genetic and Environmental Influence*, edited by R. Cancro. New York: Grune and Stratton.

_____. 1971b. *Abilities: Their Structure, Growth, and Action*. Boston: Houghton Mifflin.

Davis, James A. and Tom W. Smith. 1992. *The NORC General Social Survey: A User's Guide*. SAGE publications.

Firebaugh, Glenn. 1989. "Methods for Estimating Cohort Replacement Effects." *Sociological Methodology*, Vol. 19, pp. 243-262.

Flynn, J. R. 1994. "IQ gain over time." Pp. 617-623 in *Encyclopedia of Human Intelligence*, edited by R. J. Sternberg, S. J. Ceci, J. Horn, J. Matarazzo, and S. Scarr. New York: Macmillan.

Glenn, Norvall. 1999. "Further Discussion of the Evidence for an Intercohort Decline in Education-Adjusted Vocabulary." *American Sociological Review*, Vol. 64, pp. 267-271.

Grissmer, David, Ann Flanagan, and Stephanie Williamson. 1998. "Why Did the Black-White Score Gap Narrow in the 1970s and 1980s?" Pp.182-226 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington, DC: Brookings Institution Press.

Hauser, Robert M. 1998. "Trends in Black-White Test-Score Differentials: I. Uses and Misuses of NAEP/SAT Data." Pp. 219-249 in *The Rising Curve: Long Term Gains in IQ and Related Measures*, edited by U. Neisser. Washington, DC: American Psychological Association.

Hauser, Robert M. and Min-Hsiung Huang. 1996. "Trends in Black-White Test-Score Differentials." Discussion Paper 1110-96. Madison, Wisconsin: University of Wisconsin-Madison. Institute for Research on Poverty.

_____. 1997. "Verbal Ability and Socioeconomic Success: A Trend Analysis." *Social Science Research*, Vol. 26, pp. 331-376.

Hayes, Donald, Loreen Wolfer, and Michael Wolfe. 1996. "Schoolbook Simplification and Its Relation to the Decline in SAT-Verbal Scores." *American Educational Research Journal*, Vol. 33, pp. 489-508.

Hedges, Larry V. and Amy Nowell. 1999. "Changes in the Black-White Gap in Achievement Test Scores." *Sociology of Education*, Vol. 72, pp. 111-135.

Herrnstein, Richard J. and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Huang, Min-Hsiung and Robert M. Hauser. 1996. "Trends in Black-White Test-Score Differentials: II. The WORD-SUM Vocabulary Test." CDE Working Paper 96-30.

Madison, Wisconsin: University of Wisconsin-Madison. Center for Demography and Ecology. November.

Huang, Min-Hsiung and Robert M. Hauser. 1998. "Trends in Black-White Test-Score Differentials: The WORDSUM Vocabulary Test." Pp. 303-332 in *The Rising Curve: Long Term Gains in IQ and Related Measures*, edited by U. Neisser. Washington, DC: American Psychological Association.

Jencks, Christopher and Meredith Phillips. 1998. "The Black-White Test Score Gap: An Introduction." Pp. 1-51 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington, DC: Brookings Institution Press.

Jensen, Arthur. 1980. *Bias in Mental Testing*. New York: Free Press.

Johnson, William R. and Derek Neal. 1998. "Basic Skills and the Black-White Earnings Gap." Pp.480-497 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington, DC: Brookings Institution Press.

Lynn, Richard. 1998. "Has the Black-White Intelligence Difference in the United States Been Narrowing Over Time?" *Personality and Individual Differences*, Vol. 25, pp. 999-1002.

Maddala, G. S. 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.

Neisser, Ulric. 1998. *The Rising Curve: Long Term Gains in IQ and Related Measures*. Washington, DC: American

Psychological Association.

Neisser *et al*. 1996. "Intelligence: Knowns and Unknowns." *American Psychologist*, Vol. 51, No. 2, pp. 77-101.

Ryder, Norman B. 1968. "Cohort Analysis." *International Encyclopedia of the Social Sciences*, Vol. 2, pp. 546-550.

Schaie, K. Warner. 1996. *Intellectual Development in Adulthood: The Seattle Longitudinal Study*. New York: Cambridge University Press.

Shryock, Henry S., Jacob S. Siegel, and Associates. 1976. *The Methods and Materials of Demography*. Academic Press.

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

Wilson, James A. and Walter R. Gove. 1999a. "The Intercohort Decline in Verbal Ability: Does it Exist?" *American Sociological Review*, Vol. 64, pp. 253-266.

_____. 1999b. "The Age-Period-Cohort Conundrum and Verbal Ability: Empirical Relationships and Their Interpretation." *American Sociological Review*, Vol. 64, pp. 287-302.

Winship, C. and L. Radbill. 1994. "Sampling Weights and Regression Analysis." *Sociological Methods and Research*, Vol. 23, pp. 230-257.

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

Zeidner, J. and A. J. Drucker 1983. *Behavioral Science in the Army: A Corporate History of the Army Research Institute*. U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

美國黑人與白人在語文測驗表現逐漸接近的趨勢：出生年次分析與時期分析

黃敏雄、勞勃．豪斯

摘　要

我們使用美國「社會變遷調查」(General Social Survey) 一九七四到一九九八年的資料，擴充及修改之前發表的研究分析並得到一致的研究結果。由於美國黑人社經背景及教育機會的改善，美國黑人與美國白人在語文能力上明顯有愈來愈接近的趨勢。我們強調出生年次分析才能提供正確的黑人白人語文能力差異趨勢。由於趨勢的改變主要是循著出生年次的時間演進，以資料調查年份為時間單位的趨勢分析模糊並低估黑人白人語文能力差異逐年接近的顯著性。儘管如此，當我們適切地處理資料，以資料調查年份為時間點的趨勢分析仍然指出黑人白人在語文測驗的表現上，從一九七四到一九九八年有明顯逐年接近的趨勢。

關鍵詞： 黑人白人測驗表現差異、出生年次分析、時期分析、語文能力、WORDSUM