Lessons from the Bell Curve

Author(s): James J. Heckman

Source: *Journal of Political Economy*, Oct., 1995, Vol. 103, No. 5 (Oct., 1995), pp. 1091-1120

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/2138756

Lessons from the Bell Curve

James J. Heckman

University of Chicago and American Bar Foundation

This paper examines the argument presented in *The Bell Curve*. A central argument is that one factor—*g*—accounts for correlation across test scores and performance in society. Another central argument is that *g* cannot be manipulated. These arguments are combined to claim that social policies designed to improve social performance cannot be effective. A reanalysis of the evidence contradicts this story. The factors that explain wages receive different weights than the factors that explain test scores. More than *g* is required to explain either. Other factors besides *g* contribute to social performance, and they can be manipulated.

It has been said that there are no more intellectuals (Jacoby 1987). What is meant by this is that there are few generalists inside or outside of academia who can effectively synthesize diverse scholarly literatures to create a coherent vision of society.

Richard Herrnstein and Charles Murray attempt such a vision. The *Bell Curve* is a well-written synthesis of several social science literatures that presents a bold account of disturbing forces operating in contemporary American society. The book explores the sources of human differences and their consequences for social policy, mounting a challenge to widely held beliefs about the fundamental equality and mal-

I have benefited from numerous comments received from Sandy Jencks, Derek Neal, and Jeffrey Smith as well as helpful remarks from Gary Becker, Jennifer Boobar, John Cawley, David Donoho, Stephen Feinberg, Sam Peltzman, and Tomas Philipson. I have benefited from discussions at Carnegie Mellon, Haverford College, and the seminar on Meritocracy and Inequality at the Harris School supported by the Russell Sage Foundation. Part of this review refines and extends a more popular discussion presented elsewhere (Heckman 1995). This research was supported by grants from the Spencer and Mellon Foundations.

[*Journal of Political Economy*, 1995, vol. 103, no. 5]

leability of human beings that have been the cornerstone of American social policy since the 1960s.

Despite appearances to the contrary, this book is not a scholarly work in the conventional sense of the term. In places, its use of both primary and secondary evidence leaves much to be desired. Nonetheless, the book is worth reading because it offers a fresh—if sometimes flawed—perspective on topics that are rarely openly discussed in contemporary academic discourse and are usually settled by implicit agreement among persons with shared political and social values.

Some of the topics discussed in this book—such as the role of heritability and IQ in explaining group differences in social achievement—have become taboo in "respectable" academic discourse. Persons who assign a significant role to IQ in determining performance in the labor market or who suggest that genetic factors may account for the persistence of racial differences in social performance are routinely dismissed as racist bigots. That was the fate of Jensen and Herrnstein more than two decades ago (Herrnstein [1973] discusses this controversy), and Herrnstein and Murray have suffered the same fate in the popular press. Within academia, it has been folly for scholars in pursuit of peer-reviewed publication and peer-reviewed funding to critically examine these issues. By raising these and other forbidden topics in a best-selling book, Herrnstein and Murray perform a valuable service for the academic community. They have written a bull's-eye for academics to fire at and have written it in such a way that they are certain to draw fire.

When Sigmund Freud was asked by a young scholar how to become a successful academic, he said "exaggerate."[1] By exaggerating their evidence, Herrnstein and Murray have attracted the attention of academics and others. They deserve credit for opening the discussion of issues that have been off-limits for too long. They launch but by no means resolve the discussion of the sources and consequences of inequality. By making bold statements that cry out for more qualification, they have created a market for scholars to clarify their argument. Their book is certain to provoke a scholarly response to each of the major points raised by their work and thereby to promote the serious study of the sources and consequences of human inequality.

This review article places the *Bell Curve* in the context of current discussions of inequality and summarizes the main arguments in the book. It goes on to discuss the quality of the authors' evidence and to suggest what lessons should be learned from their study.

[1] Story relayed by Robert Cooter via John Donohue about a conversation overheard by David Daube of Berkeley at a conference in London that Daube attended with Freud.

I. Summary of the Main Argument of the Book

The focus of this book is the growth of economic and social inequality in American society. Like other writers on this topic, Herrnstein and Murray attribute the growth in inequality to the rise in the price of skilled labor relative to unskilled labor. Unlike the other authors, they equate skill with IQ. Partitioning of society by skill is synonymous in their view with a partitioning of society into high-IQ and low-IQ classes. Their equation of IQ with skill and their implicit assumption that only one skill is valued by society are central to their argument.

Herrnstein and Murray agree with authors such as Labor Secretary Robert Reich (1991) and take it as inevitable that the forces of technology and international trade will continue to operate to increase the economic return to skilled labor. Reich believes that the appropriate response to this phenomenon is to make everyone in America more skilled, although he presents no evidence about the costs of effecting such a transformation. One estimate suggests that it would take more than $1.7 trillion of investment in 1989 dollars to restore the wage losses experienced by unskilled males since 1979 (Heckman, Roselius, and Smith 1994).

Herrnstein and Murray would argue that the dollar cost figure is far higher. They view skill as genetically determined—at least in large part—and are much less optimistic than Reich about the ability of society to upgrade human skills. None of these authors entertains the possibility that the forces of trade and technology that give rise to a skill premium might be arrested. Current trends are interpreted as inevitable forces despite well-documented reversals of skill differentials in wages over long periods of time documented by Goldin and Margo (1992).

The growing premium accorded to skill in the labor market is interpreted by Herrnstein and Murray as vindicating forecasts by Herrnstein in *IQ in the Meritocracy* (1973). In his earlier work, he claimed that American society was becoming stratified on the basis of intelligence. At the time of publication of this book, the labor market premium for skills was declining, not rising, and Herrnstein did not discuss wages, but rather focused on the relationship between IQ and college admissions and occupational attainment. If cognitive stratification was at work then, it took a different form than it does now, suggesting that a richer story might be necessary to justify the economic history of the past 40 years. Inequality is a multidimensional problem, but Herrnstein and Murray consider only a unidimensional version of it. They hold that a single factor of intelligence accounts for inequality in modern society.

A. *Part 1*

Part 1 of the *Bell Curve* updates Herrnstein's earlier work and claims that American society has become more stratified on the basis of intelligence than it was even one generation ago. The authors of the *Bell Curve* correctly note that college admissions policies have led to greater stratification in colleges based on intelligence. They also claim that workplaces have become more stratified on the basis of IQ, but most of their evidence pertains to stratification by education and skill and not by IQ. Sociologists have documented that the educational attainments of husbands and wives have become more positively correlated in recent years (Mare 1991). Here, as throughout the book, Herrnstein and Murray equate IQ and skill (education) despite the fact that the correlation between IQ and skills is not the same for all skills and is certainly much less than one. Most of the evidence they offer on cognitive partitioning pertains to sorting on measured skill levels.

An increase in cognitive stratification in schools and the workplace could be the consequence of a successful realization of the meritocratic principle that access to academic and social institutions should be based on individual ability and not on social origins. A growth in the application of the meritocratic principle would imply that, when one conditions on ability, the effects of family background should have become weaker in recent decades. The authors present no evidence that this has occurred. The available evidence indicates strong effects of family background on educational attainment even in the 1980s (Grusky and DiPrete 1990). The pattern of family influence on transitions across grades has been stable throughout most of the century (Mare 1980; Cameron and Heckman 1993*a*). In sociological studies of status attainment, the weights placed on family income and family background have decreased and the weight placed on education has increased (Grusky and DiPrete 1990). The coefficient of education in earnings equations has increased in the past 15 years. Such evidence is inherently ambiguous about the growth of meritocracy because it does not control for ability. Taber (1994) documents that the rise in the return to education is not solely a consequence of the rise in the payment to the measure of ability used by Herrnstein and Murray. Controlling for their measure of ability does not budge the trend in the increase in the return to skill.

The persistence of dependence of educational outcomes on family origins may arise from either genetic or environmental factors. A rise in the meritocracy may make social origins more or less important depending on the strength of family genetic transmission mechanisms, which are not known with any precision.

Evidence of growing economic inequality among skill classes is confirmed in many recent studies. The source of the skill differences is much less clear. It is at this point in their argument that Herrnstein and Murray part company with conventional social science and venture into hard-core psychometrics, human genetics, and heritability. They develop two points that are central to their argument.

First, they appeal to a psychometric tradition dating back to Spearman (1904) and assert that *one* latent ability factor called *g* can explain the high degree of correlation typically found among scores of different achievement and ability tests administered to the same individuals. The term *g* stands for general ability, or intelligence. By no means is there agreement in the psychometric community that one factor is sufficient to "explain" the data—in the sense of explaining the fit of a matrix of correlations of test scores[2]—or that there is just one type of intelligence (see, e.g., Sternberg 1985, 1988; Carroll 1993; Hunt 1994). Below I present evidence about *g* using the same data used by Herrnstein and Murray.

The notion of *g* has been revived in recent studies that predict performance in military vocational training schools (see, e.g., Ree and Earles 1992) and employee supervisor ratings (Jensen 1986). This literature claims that one weighted average of test scores explains "most" of the correlation pattern found among tests. It is further claimed that *g* is all that is needed to predict performance in workplaces and in military training programs, although statistical tests indicate that additional factors are statistically significant in performance prediction equations (see, e.g., Ree and Earles 1992). The *g* factor is viewed as general intelligence or IQ that explains performance in the workplace and also performance on intelligence and achievement exams. However, Herrnstein and Murray do not note the low R^2 in such prediction equations: rarely are they as high as 20 percent in predicting outcomes in military training programs or as high as 30–40 percent in predicting supervisor performance ratings.[3] In log wage equations, the R^2 is no more than 10–12 percent for white males and is as low as 7 percent for hourly wage equations when test scores alone are included in the equations (see Cawley et al. 1995).

Second, drawing solely on published sources, Herrnstein and Murray claim that *g* and hence the consequences of *g* are heritable, with heritability in the range of 40–80 percent. This says that 40–80 per-

[2] The usual measure of fit is the trace of the correlation matrix of test scores. Formally, *g* is the value of all the test scores weighted by the eigenvector associated with the largest eigenvalue of the correlation matrix of test scores.

[3] These R^2 statistics are called "validities" in the psychometric literature.

cent of the variance of IQ *measured around population means* is attributable to biological genetic influences. The authors qualify this claim with the usual warnings about the possibility of interaction effects: that at different levels of the environment, genetic factors may play different roles in accounting for variability in outcomes among individuals. In the extreme, if there were no differences in the environment facing persons, all variability would have to be due to genetic factors. This observation accounts for the fact that the heritability of IQ in egalitarian Norway is higher than it is in the United States.

The authors ask us to accept that a substantial component of the variance in outcomes is attributable to genetic factors. At certain places in the text, they write as though the heritability factor is as large as 100 percent.

In developing these points, Herrnstein and Murray set the record straight about influential intellectual gossip directed against psychometric testing and the measurement of heritability. A literature emerged in the late 1970s and early 1980s that attempted to discredit testing and the importance of heritability in defense of the prevailing environmentalist point of view. The books by Kamin (1974) and Gould (1981) are good examples of this genre. They attempt to discredit psychometric research by linking it to crude racist theories and policies, blaming the misapplication of psychometrics on the psychometricians. Both authors rely heavily on innuendo and arguments based on guilt by association, and neither acknowledges the well-established facts that IQ has both predictive power in the labor market and a substantial heritable component. Herrnstein and Murray show that much of the popular literature is misleading in its treatment of modern psychometrics.

In this part of their book and in part 3, Herrnstein and Murray discuss the well-known fact that psychometric tests of worker performance are biased *in favor* of blacks and other minorities; that is, at a given level of a test score, regressions of productivity on test scores based on pooled samples of blacks and whites predict a higher level of job performance for blacks than for whites. A more balanced assessment, however, would also point out that if such fair test scores are used in conjunction with a threshold below which persons are not hired, high-productivity blacks tend to be eliminated at a greater rate than high-productivity whites (Hartigan and Wigdor 1989).

Claims that tests are culturally biased and hence should be discarded in making job and schooling assignments have no basis in fact. The claim that Sir Cyril Burt forged his data in his influential studies of heritability is noted to be false. His correlation patterns have been duplicated in later studies on independent samples (see, e.g., Joynson 1992). The fact that IQ tests and notions of heritability have been

misused by some zealots does not deny that IQ predicts productivity, albeit with an R^2 much less than one; that heritability explains much (40–80 percent) *within-group* variability; and that the best available evidence indicates that IQ and achievement tests are not culturally biased.

B. *Part 2*

Part 2 of the *Bell Curve* presents original empirical research on the relationship between the authors' measure of IQ and performance in society. Herrnstein and Murray exploit a large (12,000-person) nationally representative sample, the National Longitudinal Survey of Youth (NLSY) 14–24 years old in 1979 that is still ongoing. These data are uniquely valuable for establishing a relationship between ability and outcomes because of the wealth of information on labor market and social outcomes of participants in the survey. These data contain scores on 10 tests developed by the military to predict performance in armed forces training programs (the Armed Services Vocational Aptitude Battery [ASVAB]). The battery of tests consists mainly of unspeeded achievement tests designed to measure knowledge of vocabulary, basic science, arithmetic operations, and the like, although there are two speeded tests of numerical and coding operations. There are also tests of specific vocational skills. The Armed Forces Qualifying Test (AFQT) used by Herrnstein and Murray as their measure of IQ is assembled from a simple unweighted average of four of the achievement tests. The AFQT is the most commonly used combination of the tests used by the military to predict performance. This test is not the same as the g that can be extracted from the full battery of 10 tests available on the survey, although it is highly correlated with it (see Cawley et al. 1995). While proclaiming the virtues of g, Herrnstein and Murray do not actually use it in their empirical analyses.

A variety of interesting empirical associations are established in this portion of the book. Although the authors issue the standard warning that correlation does not imply causation, throughout the book these correlations are given an implicit causal interpretation, as is common in much empirical research in social science. This problem is especially serious in their analysis of the relationship between AFQT and education, as I note below.

A number of interesting empirical regularities emerge from Herrnstein and Murray's logit regression analyses of the NLSY data. Low-AFQT (IQ) persons are more likely to be in poverty, drop out of school, be unemployed or out of the labor force, be on welfare, be bad parents, commit crimes, and fail to participate in political

activity than are high-AFQT (IQ) persons. This relationship holds even after they adjust for a scalar measure of socioeconomic background discussed below.

These exercises show that AFQT—IQ—plays a powerful role in generating differences in a variety of socially important behaviors. The authors develop a novel way of demonstrating this result that does not depend on the units in which IQ or socioeconomic status is measured. They compare the change in the probability of an outcome when IQ ranges over 95 percent of the population (from the top 2.5 percent to the bottom 2.5 percent) with the change in the probability of an outcome when socioeconomic status varies over the same 95 percent range. The variation in each measure is evaluated at the mean (or median) of the other measure. (It would be more convincing if this exercise had been conducted using multiple points of evaluation and not just at the medians, as the authors do.) Using logits ($\ln[P/(1 - P)]$, where P is the probability of the event being studied), the authors obtain the same results. For either the logits or the probabilities of virtually all their outcome measures, a 95 percent change in the value of the socioeconomic status variable has much less effect in changing outcomes than a 95 percent change in IQ. In this sense, IQ is the more powerful variable. Interpreted causally, a variation from the top 2.5 percentile of IQ to the bottom 2.5 percentile of IQ affects behavior more than a comparable change in socioeconomic status.[4] Below I discuss difficulties with interpreting their measure of relative importance.

This analysis is conducted only for whites. By proceeding in this way, the authors establish the importance of IQ in accounting for individual differences without getting into the controversial issue of racial bias in IQ tests.

C. *Part 3*

Part 3 is the most controversial portion of the book, if only because it discusses black-white differences in IQ. Herrnstein and Murray analyze the sources of ethnic differences in social outcomes and the role of their measure of IQ in accounting for these differences. They cite numerous scholarly studies that refute critics of IQ and aptitude tests who claim that they are racially or culturally biased and unrelated to true productivity in schools or the workplace, and that differences in test scores are not predictive of differences in performance

[4] The authors' measure is thus invariant to monotonic changes in the scaling of the IQ or socioeconomic status variable, but obviously depends, in general, on the point of evaluation of the variable being held constant.

in society. They demonstrate convincingly that psychometric tests predict productivity, even if not perfectly.

The authors devote considerable attention to genetics and the inconclusive debate about the role of genes and cultural and environmental influences in explaining ethnic differences in IQ and social performance. Their discussion is remarkably evenhanded and does not make a decisive case against the dominance of environmental factors, although the authors clearly wish their readers to share their belief that genetic factors are more important. One of their primary exhibits in support of that belief is that whites and blacks differ more on tests most correlated with *g* and less on those least correlated with *g*. They discuss the important problem of interaction between genes and environment in defining unique contributions of either to intelligence. They make the correct observation "that a trait is genetically transmitted in individuals does not mean that group differences are also genetic in origin" (p. 298).

Herrnstein and Murray do not adequately develop the importance of the difference between means and variances that is crucial to understanding the evidence from the heritability literature (see, e.g., Cavelli-Sforza and Bodmer 1971, p. 794). Compare two populations, *B* and *W*. Suppose that all persons in *B* are at level X_B of an environmental variable. All persons in *W* are at level X_W of the environmental variable. Let $\mu_B(X_B)$ be the *B* mean and $\mu_W(X_W)$ be the *W* mean.

Outcomes Y_B and Y_w may be written as

$$Y_B = \mu_B(X_B) + \epsilon_b,$$

$$Y_w = \mu_w(X_w) + \epsilon_w,$$

$$E(\epsilon_b) = E(\epsilon_w) = 0,$$

where ϵ_b and ϵ_w are the residual deviations from population means. Within the *W* population or the *B* population, *none* of the variance can be explained by levels of the *X* factor because they are common to all persons. Across populations, the mean difference depends on the level of the *X* factor. Even if *X* varies between the two groups, the total contribution of *X* to explaining between-group differences is understated in standard heritability calculations. In the language of the analysis of variance, heritability measures the genetic contribution to within-group variance. Heritability calculations tell us nothing about the contribution of the environment to between-group differences or the evolution of the group mean over time. Height is a highly heritable characteristic. Yet the height of the Japanese population has increased by 5–7 inches in one generation because of widespread improvement in diet.

Herrnstein and Murray discuss the "Flynn effect": that IQ scores

drift upward over time. Within one generation, the increase in IQ within any population has the magnitude of the gap in IQ between blacks and whites. The authors' favored explanation for this phenomenon is that egalitarian interventions have raised the measured IQ of the lower tail of the distribution. They ignore this evidence in part 4 when they discuss the ineffectiveness of social interventions.

Herrnstein and Murray present empirical work that substantiates the role of their measure of IQ in accounting for a considerable portion of ethnic differences in socioeconomic outcomes. They demonstrate that social pathology is concentrated among low-IQ persons.

Turning to traditional eugenic themes, the authors discuss the higher rate of reproduction and immigration of lower-IQ groups. They also discuss the consequences of this phenomenon for the American gene pool. They claim that partly as a consequence of compositional change, the average IQ is declining in the United States. However, in contradiction to their own evidence on the Flynn effect, they also claim that throughout most of this century, IQ has been decreasing within groups as well.

The argument in this part of the book is much less clearly stated than the argument in the earlier parts of the book. Despite the authors' emphasis on genetics, they present much powerful evidence in support of an important role for environmental factors in accounting for racial and ethnic differences in IQ. Their argument about dysgenic development does not satisfactorily account for their own evidence on the Flynn effect.

D. Part 4

Part 4 discusses the consequences of the preceding analyses for social policy and social trends in the United States. It consists of two separate sections. The authors note the tensions inherent in a society based on the Jeffersonian principle that "all men are created equal" when social and economic forces produce profound differences in social performance among persons. Strictly speaking, Thomas Jefferson's phrase applies to moral sensibility, which he held to be common to all persons, and not to equality in intelligence or market ability. Herrnstein and Murray argue that physical and social separation of groups undermines the functioning of cooperation and promotes factions in a democratic society. This is a view they share with Reich (1991) and Kaus (1992).

The first section builds on the first three parts of the book and discusses the implications of the authors' findings for social policy. Herrnstein and Murray present a pessimistic summary of efforts to raise cognitive ability through social programs. This review of the

ineffectiveness of most social programs harkens back to Murray's *Losing Ground* (1984), except now the cognitive limitations of individuals are stressed rather than the perverse incentives created by those programs.

There is no necessary inconsistency between these two positions. This book stresses the location of the budget constraint facing individuals and the inability of cognitively weak individuals to make wise choices. As pointed out by Becker (1962), responses to incentives are generated by the budget constraint and not by the utility function, so that even cognitively feeble persons respond to shifts in budget constraints or nonmarket incentives induced by social programs. Even idiots cannot spend beyond their means, and it is the budget constraint and not the utility function that produces downward-sloping demand curves and responses to incentives.

Herrnstein and Murray discuss the "dumbing-down" of American public education and the shift in educational expenditures away from gifted children. In the name of promoting equality, educational policies have taken resources away from the able and given them to the less able. Special education has come to mean integrating the "learning impaired" into the classroom and not promoting the education of the brightest.

Herrnstein and Murray demonstrate that disparities in intelligence and abilities among ethnic groups and policies of equality of opportunity produce demographic disparities in performance in schools and the workplace. Such disparities often attract the attention of civil rights officials and affirmative action officers. The authors survey a large literature (well summarized in Hartigan and Wigdor [1989]) that establishes that employment and aptitude tests predict productivity and are biased *in favor* of minorities. They go on to claim that prohibitions against the use of tests impair economic productivity. They rely on the highly controversial work of Hunter and Schmidt (1982) to produce an estimate of the cost of not using tests to select workers. Most scholars reject the Hunter-Schmidt estimates as grossly inflated (see, e.g., Hartigan and Wigdor 1989).

The press has attacked this section of the book, as well as the section on racial differences in IQ, as racist in tone and content. Contrary to impressions in the popular press, the authors advocate the nonracist policy of treating persons as individuals rather than as members of racial groups. They make the apparently controversial observation that racial disparities in performance are not solely attributable to discriminatory policies by firms or institutions and that equality of opportunity does not imply equality of outcomes.

The final chapters of the book abandon the earlier empirical focus. The penultimate chapter raises the possibility that the cognitive elite

could become increasingly isolated from society at large. Similar fears have been expressed by Robert Reich in his *Work of Nations* or Mickey Kaus in *The End of Equality*. In one scenario, the affluent and the cognitive elite merge interests to form a ruling class. Indeed, many of the cognitive elite are affluent as a result of the rise in the economic return to skill. As the forces of technology unfold relentlessly, the less able become unemployable burdens on society at large. This portion of the book reads like a piece of science fiction, with about as much factual support.

The final chapter of the book turns to the paradise envisioned in Murray's *In Pursuit* (1988) as a solution to the problem of increasing stratification by IQ in society at large. That work extols the virtues of small-town America and envisions a "communitarian" order in which "places" are found for everyone much as small-town America found places for the village idiot, the village drunk, and the spinster schoolteacher. Herrnstein and Murray express strong faith in the power of local communities to solve the social problems discussed in their book, but they do not discuss the specific mechanisms through which this solution would be accomplished. Essential features of their communitarian order are simplicity of rules and clear definitions of right and wrong that are so well formulated that even the cognitively feeble can understand them. With these rules, they hope to restore Jefferson's moral sensibility to the population at large.

This book raises many interesting questions. The authors' challenge to contemporary assumptions about the malleability of human beings and the relative importance of environmental factors in producing differences in individual outcomes is courageous and long overdue. It is unfortunate that the authors do not present a more effective challenge.

II. Why the Book Fails

The book fails for five main reasons.

1. The central premise of this book is the empirically incorrect claim that a single factor—*g* or IQ—that explains linear correlations among test scores is primarily responsible for differences in individual performance in society at large. Below I demonstrate that a single factor can *always* be constructed that "explains" all correlations in responses to a test or correlations in scores across a battery of tests, but in general this *g* is not constructed by conventional linear methods. There is much evidence that more than one factor—as conventionally measured—is required to explain conventional correlation matrices among test scores. Herrnstein and Murray's measure of IQ is not the same as the *g* that can be extracted from test scores available

in their data set. They do not emphasize how little of the variation in social outcomes is explained by AFQT or *g*. There is considerable room for factors other than their measure of ability to explain wages and other social outcomes.

2. In their empirical work, the authors assume that AFQT is a measure of immutable native intelligence. In fact, AFQT is an achievement test that can be manipulated by educational interventions. Achievement tests embody environmental influences: AFQT scores rise with age and parental socioeconomic status. A person's AFQT score is not an immutable characteristic beyond environmental manipulation.

3. The authors do not perform the cost-benefit analyses needed to evaluate alternative social policies for raising labor market and social skills. Their implicit assumption of an immutable *g* that is all-powerful in determining social outcomes leads them to disregard a lot of evidence that a variety of relevant labor market and social skills can be improved, even though efforts to boost IQ substantially are notoriously unsuccessful.

4. The authors present no new evidence on the heritability of IQ or other socially productive characteristics. Instead, they demonstrate that IQ is more predictive of differences in social performance than a crude measure of parental environmental influences. This comparison is misleading. It fails to recognize the crudity of their environmental measures and the environmental component that is built into their measure of IQ, which biases the evidence in favor of their position. Moreover, the comparison as they present it is intrinsically meaningless.

5. Finally, the authors' forecast of social trends is pure speculation that does not flow from the analysis presented in their book. Most of the social policy recommendations have an ad hoc flavor to them and do not depend on the analysis that precedes them. The appeal to Murray's version of communitarianism as a solution to the emerging problem of inequality among persons is a deus ex machina flight of fancy that is not credibly justified.

I now amplify each of these points in turn, presenting fresh empirical results developed jointly with students at the University of Chicago in the Winter 1995 course Economics 311.

A. *Is There a Single* g*? The Diversity of Human Skills*

Nothing central to the case for recognizing diversity in human abilities hinges on the issue of whether there is one ability or whether there are multiple abilities, as much econometric and psychometric research, and the data analyzed by the authors, indicates is the case.

Indeed, the existence of multiple abilities bolsters the empirical case for heterogeneity in ability as an important fact of social and economic life.

Yet Herrnstein and Murray cling to g and devote long sections of their book to its defense. For them, g has three essential characteristics: (*a*) it explains correlations among test scores, (*b*) it explains differences in achievement across persons, and (*c*) it is an immutable characteristic of a person that can be changed only at a prohibitive cost. Their staunch defense of g is linked to its traditional use in studies of heredity, although heredity is not intrinsically a scalar concept. In principle, heredity could operate on vectors of ability, although the problem of estimating heritability would be correspondingly complicated[5] (see the discussion in Carroll [1993]).

Herrnstein and Murray offer no fresh empirical evidence on g but refer to an existing published literature in psychometrics: primarily research by Jensen (1986) and a series of papers by Malcolm Ree and his associates (see, e.g., Ree and Earles 1992; Olea and Ree 1994; Ree, Earles, and Teachout 1994). The research by Ree and Earles uses ASVAB scores of the sort available in the data analyzed by Herrnstein and Murray. Ree and Earles study the intercorrelation among the ASVAB scores and the power of g in predicting success in vocational schools operated by the armed forces. One factor appears to be dominant in explaining the ASVAB correlation matrix.[6] But the second component is not negligible.

The g factor explains 17 percent of the variance in pass rates in military training schools. A second factor raises R^2 to 18 percent. While the second factor is strongly statistically significant (samples include 80,000 recruits), it is striking that the addition of a second score generated from the second principal component of test scores does so little to improve the fit of the prediction equation. Of course it is equally striking that the R^2 with one or two factors is only 17–18 percent. The explanatory power of g is far from perfect.

The authors are inconsistent about their use of R^2 and conventional statistical significance tests to justify inclusion of variables in estimated models. They reject using R^2 as a model selection criterion in favor of conventional significance testing procedures (p. 594), but by this standard their major source of evidence on g (e.g., the Ree and Earles studies) indicates that more than one "factor" or ability is necessary to explain correlations among the ASVAB scores. This methodological inconsistency pervades their book.

[5] Traditional analyses of parent-child IQ scores would not identify a vector of skills. Multiple measures and multiple factor analysis would be required.

[6] Ree and Earles demonstrate that g is highly correlated with conventional factor scores.

The *g* factor is an artifact of linear correlation analysis. A theorem of Suppes and Zanotti (1981) informs us that for any vector of test scores from an achievement test, it is possible to construct a scalar latent factor such that, conditional on the factor, test scores are independent.[7] The *g* factor exists for any vector of binary, finite-valued, or countably valued random variables. The *g* of conventional psychometrics is a product of mathematical conventions in factor analysis. A *g* also exists to account for correlations *among* test scores. That is a mathematical theorem of no behavioral consequence for psychometrics or for finance, another field addicted to factor models. The value of *g* in predicting behavior is the real test of its importance. There is much evidence that it has predictive power.

The psychometric community is far from unanimous in endorsing *g*. In a major reanalysis of the major data sets of psychometrics, Carroll (1993) concludes that there are three important sorts of cognitive abilities corresponding to fluid intelligence (ability to solve problems quickly), crystallized intelligence (the ability to draw on old solutions to address new problems), and spatial and mechanical ability. Sternberg (1985, 1988) documents the commonsense point of view that there are many market—and performance—relevant skills. Personality, motivation, and drive are all valued by society. Firms screen and hire on all these characteristics. Sternberg supplements his academic income by designing personality and motivation tests as well as tests for specific abilities that are valued in different occupations.

None of this evidence denies that *g* plays a nontrivial role in explaining social outcomes. However, it is not monolithic. For the ASVAB scores in the NLSY used by Herrnstein and Murray, one factor is dominant (see table 1). The *g* factor accounts for 55–71 percent of the variance in test scores across demographic groups. It is a better predictor for men than for women. For women the subsidiary factors are generally more important, but the differences in the weights between the sex groups are small. The first factor (*g*) heavily weights the achievement test component of ASVAB. The second factor weights the speeded operations that are usually thought to be good measures of "fluid intelligence" or "problem-solving ability."

[7] Suppes and Zanotti (1981) show that if $\mathbf{D} = (d_1, \ldots, d_T)$ is a vector of T binary random variables with density $f(\mathbf{D})$, then there always exists a factor g such that

$$f(\mathbf{D}|g) = \prod_{i=1}^{t} f(d_i|g),$$

so that g plays the role of a single factor in conventional factor analysis. Standard probability arguments can be used to extend the theorem to countably valued random variables (e.g., success proportions on exams) and hence to approximate continuous variables arbitrarily well (see, e.g., Holland and Rosenbaum 1986).

TABLE 1

PROPORTION OF VARIANCE IN TEST SCORES ATTRIBUTABLE TO PRINCIPAL FACTORS

Demographic Group	g_1	g_2	g_3	g_4	g_5	Cumulative (%)
Black females (N = 1,504)	.550	.096	.070	.063	.060	.842
Black males (N = 1,524)	.640	.085	.060	.050	.035	.868
Hispanic females (N = 933)	.650	.079	.054	.043	.039	.866
Hispanic males (N = 910)	.706	.081	.052	.037	.028	.904
White females (N = 3,502)	.579	.108	.068	.058	.042	.856
White males (N = 3,541)	.638	.113	.059	.046	.031	.889

SOURCE.—Cawley et al. (1995).

NOTE.—If Σ is the correlation matrix of ASVAB test scores, the conventional measure of the variance used is the trace tr $\Sigma = \Sigma_{i=1}^{I} \sigma_{ii}$. g_1 is the same as g in the text.

The weighting placed on various "abilities" depends critically on the composition of the tests used to measure ability. If the achievement tests used to define AFQT are deleted from the ASVAB scores, the g loaded scores become the speeded tests.

The g factor extracted for all demographic groups is remarkably similar. There is less similarity across demographic groups in the other factors. This evidence is supportive of a single dominant factor as a determinant of test scores (see the evidence in Cawley et al. [1995]). The g measure is not the same as the AFQT measure used by Herrnstein and Murray, although it is strongly correlated with it (R^2 = .97). Arguably by not using g in their analyses of social performance, they bias downward their estimate of g's effect on social outcomes, but the difference between AFQT and g as predictors is slight.

It is more interesting to consider how well g predicts outcomes in society. Wages are a variable of great interest to economists as a measure of performance in society. For each major demographic and ethnic group, table 2 presents simple Pearson correlations for log hourly wages in 1991 from the NLSY survey used by Herrnstein and Murray, with g obtained from group-specific factor analyses. The table also presents correlations of log wages with the second factor (g_2), with AFQT, and with numerical operations—one of two speeded tests in the ASVAB battery.

This table exhibits the same g dominance that is found in a variety of studies cited by Herrnstein and Murray. There are several other noteworthy features of this table. First, the test scores predict female wages better than male wages. Second, g and AFQT perform about equally well in predicting wages for most demographic groups. Third, neither g nor AFQT explains all that much of the variance in log wages. The highest R^2 is less than 22 percent. A lot of variability in log wages remains unexplained. Even if measurement error is as

TABLE 2

CORRELATIONS OF LOG HOURLY WAGES WITH VARIOUS ABILITY MEASURES FOR NLSY PERSONS WITH A WAGE AND OUT OF SCHOOL, 1991

Demographic Group	AFQT	g	g_2	Numerical Operations	All Tests
Black females (N = 957)	.4527	.4536	−.0961	.3729	.4879
	(.0001)	(.0001)	(.0029)	(.0001)	(.0001)
Black males (N = 1,031)	.3449	.3510	.0432	.2774	.3711
	(.0001)	(.0001)	(.1653)	(.0001)	(.0001)
Hispanic females (N = 573)	.3137	.3268	.0128	.2927	.3787
	(.0001)	(.0001)	(.7602)	(.0001)	(.0001)
Hispanic males (N = 690)	.2725	.3056	.0227	.2722	.3743
	(.0001)	(.0001)	(.5521)	(.0001)	(.0001)
White females (N = 1,716)	.3423	.3331	.1182	.2858	.3774
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
White males (N = 1,880)	.3334	.3387	.1280	.3371	.3759
	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)

NOTE.—p-values are in parentheses.

high as 30 percent of wages (a very generous estimate; see, e.g., Bound et al. [1990]), more than half of the variability across persons is explained by factors other than g or "ability." Fourth, the second factor, g_2, is never very strong and does not predict as well as speeded numerical operations, even though it is heavily "loaded" on it. Finally, a numerical operations test, a single 5-minute test, predicts log wages better than AFQT in 1991 for white males. The score on the numerical operations test is not included in AFQT. In results not reported here, the additional gain in R^2 from using all 10 ASVAB scores to predict log wages over using g by itself is never more than 2 percent except for Hispanic females, for whom the gain is 3 percent (Cawley et al. 1995). When the conventional statistical testing procedures advocated by Herrnstein and Murray (p. 549) are used for most groups, additional scores of tests beyond g—or AFQT—are justified for inclusion in the prediction equations for log wages. For white males, as many as four ability measures are statistically significant in log wage equations. Nonetheless, it is striking that the same g that predicts test scores does such a good job of summarizing how a variety of test scores predict log wages. This evidence confirms the dominance of a single factor in explaining wages that is similar to the dominance that occurs in predicting military performance.

The fact that g, AFQT, or even the entire battery of ASVAB tests explains only a fraction of the variance in measured wages (24 percent at the most, 30 percent when one adjusts for measurement error in wages), pass rates in military training schools (17–18 percent), and supervisor ratings (40–45 percent) means that there is a lot of room

for factors not measured by psychometric tests to account for the variation in performance in a number of settings.

The seminal work of Mincer (1972) suggests two important factors: education and job experience. Cawley et al. (1995) present evidence that education and tenure on the job account for a substantial component of the variance in log wages even after measures of ability are introduced into wage equations. Table 3 presents evidence on the improvement in fit of log wage equations that arises from adding education and work experience after controlling for measures of ability. This stepwise procedure obviously exaggerates the explanatory power of the AFQT test.

Adding schooling, tenure on the most recent job, and Mincer's measure of work experience raises R^2 substantially above the level obtained from a pure psychometric specification. Education and tenure on the most recent job are always statistically significant at conventional levels. For Hispanics and whites, these variables have a substantially greater proportionate effect for males than for females, doubling R^2. Psychometric variables are strongly predictive of wages, but so are schooling and work experience.[8]

Moreover, the important factors for explaining tests are not always the important factors for predicting wages. Cawley et al. (1995) note that a one-standard-deviation increase in scores on the speeded numerical operations test raises wages substantially more than a one-standard-deviation increase in g or AFQT. Numerical operations scores often drive AFQT or g scores into statistical insignificance in regressions of log wages on experience, schooling, and ability. For whites, the coefficient on numerical operations is larger than the coefficient on AFQT or g. The g that emerges from the test score matrix is dominated by a test score that is loaded on the second component. This evidence emphasizes the discrepancy between the factors that predict test scores and the factors that predict social performance. The hypothesis of a universal g that underlies the analysis of Herrnstein and Murray does not receive much support.[9]

Analyses by researchers in the military (see, e.g., Harris, McCloy, and Statman 1995) demonstrate that work experience in a job partially compensates for initial cognitive deficits, especially for tasks that are not complex and in which there is little technical change. However, in times of rapid change, the reward to ability appears to in-

[8] Cawley et al. (1995) demonstrate that these results are invariant to arbitrary monotonic transformations of the inherently ordinal test score variables.

[9] Cawley et al. document that different ASVAB tests have different predictive power for log wages for different demographic groups. The payment for verbal skills is higher for women than for men. The reverse is true for math skills. This evidence is consistent with occupational sorting by skills but is not consistent with "universal g."

TABLE 3

INCREMENT TO R^2 OF LOG WAGE EQUATIONS FROM ADDING SCHOOLING AND WORK EXPERIENCE TO A MODEL WITH A TEST SCORE (1991 Log Hourly Wages)

	g		AFQT	
	ΔR_g^2	$(\Delta R_g^2/R_g^2) \times 100$ (%)	ΔR_{AFQT}^2	$(\Delta R_{AFQT}^2/R_{AFQT}^2) \times 100$ (%)
Black females	.1414	65.4	.1393	63.7
Black males	.1158	99.3	.1127	97.1
Hispanic females	.1191	98.7	.1243	109.5
Hispanic males	.0181	18.1	.0216	27.6
White females	.1262	112.6	.1184	98.7
White males	.0606	53.2	.0592	53.5

SOURCE.—Cawley et al. (1995).

crease. In more complex work environments and environments undergoing technical change, experience is a less perfect substitute for innate ability. Moreover, even in simple tasks, unaffected by technical change, experience never eliminates ability differentials. There is a remarkable parallelism in performance profiles in terms of experience among different cognitive classes that is strongly suggestive of Mincer's evidence on the parallelism of log wage–experience profiles across different education groups (Harris et al. 1995). At all levels of experience, more able workers retain their initial advantage over less able workers. Military studies of motivation and attitude show only weak effects of those traits on performance. At least in the military, motivation and drive are negligible contributors to productivity.

This evidence and the studies of Sternberg (1985) support the notion that g is important, but g alone does not explain social outcomes. "Crystallized intelligence" or experience also contributes to social performance. See the discussion in Hunt (1994). Cameron and Heckman (1993*b*) demonstrate that although holders of General Educational Development certificates have higher AFQT scores, they earn no higher wages than other high school dropouts with the same years of education.

B. Can g Be Manipulated?

The Flynn effect, the fact that AFQT scores rise with age, studies of transracial adoption, and the research of Herrnstein et al. (1986) all suggest that g is affected by the environment. Since g is achievement test loaded and schooling is likely to raise performance on achievement tests, there is likely to be a strong relationship between g and education. In the limit, if a brilliant person had no schooling, he or

she would be unlikely to score well on exams. The correlation between AFQT and years of schooling is high (r = .6 for white males), but AFQT may affect schooling.

A direct test of this proposition is provided by Neal and Johnson (1994), who adopt an instrumental variable strategy similar to that used by Angrist and Krueger (1991). Restrictions on the age of students entering schools cause many children born in the last quarter of a year to start school one year later than students born earlier in the same calendar year. Neal and Johnson examine grades completed and AFQT scores for students who were 16–18 when they took the test. They find within each birth year that years of schooling completed are roughly constant over the first three quarters and then drop substantially (⅓–½ year) in the fourth quarter. By using quarter of birth as an instrument for years of schooling completed, they show that an additional year of schooling raises AFQT scores for men and women by 0.22 and 0.25 standard deviations. The black-white AFQT gap could be closed by four additional years of school.

The evidence that exogenous schooling substantially raises AFQT calls into question the evidence offered by Herrnstein and Murray (app. 6). It also suggests that a substantial component of their education-AFQT correlation is due to reverse causation. It is the education that "causes" some portion of the AFQT score and not the reverse.

Neal and Johnson further document that parental education, occupation, and other environmental influences are also associated with higher IQ. This evidence is ambiguous, however, because favorable family environmental characteristics may be a consequence of high levels of parental g. If so, the association between parental environment and g may reflect the operation of heredity.

Even if an environmental effect on ability is accepted, the relevant economic question to ask about any particular intervention is whether marginal benefits exceed marginal costs. I turn to that question next.

C. *The Missing Cost-Benefit Analysis*

Before I read this book, I thought that Herrnstein and Murray would do for policies aimed at reducing inequality what Murray did for poverty policy in *Losing Ground* by documenting the ineffectiveness of many social programs, including those designed to boost the IQ of the less able. Chapter 17 of their book discusses the *mixed* evidence on the success of early childhood interventions designed to boost IQ and promote the skills that raise social performance. By no means does the evidence that they discuss rule out the possibility of boosting IQ through programs that enrich the learning environments of young children. Herrnstein's own work shows that modest gains are

possible (see, e.g., Herrnstein et al. 1986). Indeed, the authors acknowledge that there are strong indications that very intensive programs such as the Perry Preschool Program can be effective for disadvantaged populations (see Schweinhart, Barnes, and Weikart 1993). Even though this program has only modest or negligible effects on IQ, it does promote social skills and motivates students to acquire market skills. A randomized evaluation of this program that follows disadvantaged youngsters over a period of 25 years reveals that students who participate in the enriched program have greater success in life using many commonly accepted measures of that term. On average, heterogeneous programs such as Head Start are not effective in raising IQ, but the best versions of these programs boost social skills that are valued in society and promote social performance (Zigler and Muenchow 1992).

It is striking that the authors do not discuss the marginal costs and benefits of various interventions. It is in these terms that rigorous public policy discussions regarding skill enhancement programs are conducted. The authors avoid all the hard work required to make credible cost-benefit calculations for skill formation programs by claiming that there is a genetic basis for skill differences.

Evidence of a genetic component to skills has no bearing on the efficacy of any social policy. Only if the sole source of skills is genes and no offsets to genetic endowments are possible would it be acceptable to avoid the hard work of doing a cost-benefit analysis in judging the suitability of any proposed policy intervention. In that extreme case, which no one—including Herrnstein and Murray—accepts as empirically relevant, the costs of raising skills would be infinite and social programs could not be effective.

Equally obvious, knowing that all skill components are environmentally determined does not in itself justify undertaking any intervention. The relevant issue is the cost effectiveness of the intervention. The authors produce no systematic evidence on this question for any social programs designed to boost skills.

Insufficient attention is paid to the role of education in boosting earnings. Herrnstein and Murray cling to their assumption that the return to education is a return to ability. In so doing, they ignore a vast literature in social science that demonstrates that controlling for ability lowers—but by no means eliminates—the return to schooling.

The evidence on this issue is consistent across many studies. When one controls for the Herrnstein-Murray measure of ability, the returns to education sometimes fall by as much as 35 percent, but they do not go to zero. Ability and education are distinct, and both have economic rewards. On average, an extra year of schooling still increases earnings by at least a substantial 6–8 percent (Taber 1994).

There is plenty of room for social policy to eliminate earnings differentials among persons of the same ability level.

Empirical evidence by Taber (1994) and Murnane, Levy, and Willett (1995) indicates that the increase in the return to the measure of ability used by Herrnstein and Murray accounts for only a small fraction of the recent increase in the economic return to schooling. Putting their measure of ability into wage or earnings functions does not eliminate the increase in the return to schooling or work experience that has occurred in the past 15 years, nor does it eliminate the return to schooling or experience. There is a lot of evidence that market-relevant skills can be improved for all but the most poorly motivated and cognitively dysfunctional persons.

At the same time, neither the literature on schooling nor that on job training programs provides the evidence required to evaluate the cost effectiveness of education and training as tools for equalizing the earnings of persons of different ability levels. The available evidence suggests that ability—or IQ—is not a fixed trait for the young. Sustained high-intensity investments in the education of young children, including parental activities such as reading and responding to children, stimulate learning and promote education, although they do not necessarily boost IQ by very much. The available evidence suggests that such interventions stimulate motivation and social performance in the early adult years even if they do not raise IQ. There is fragmentary evidence that enriched education can be a good investment even for children of low *initial* ability because it stimulates cumulative learning processes (Moore 1986; Schweinhart et al. 1993). Similar interventions for adults for whom ability and motivation are more stable characteristics are much less effective. For low-ability adults and adults with serious motivational problems, there is little evidence that educational investments are economically justified. Wage subsidies may be a better strategy for integrating such groups into society (Phelps 1994).

D. Comparing the Role of Environment and Ability

Herrnstein and Murray examine the predictive performance of one measure of cognitive ability. They pit their ability measure against a measure of the socioeconomic status of persons when they were children. The authors intend this contrast to shed light on the relative importance of "ability" and "environment" in accounting for behavior. Given their emphasis on the genetic basis for ability, this comparison is also intended to be informative on the relative strength of genes and environment. Outcomes are shown to be much more sensitive to their measure of ability than they are to their measure of socioeco-

nomic status. Large changes in the background measure have weak effects on many outcome measures, whereas small changes in ability have large effects on the same outcome measures.

The credibility of any empirical study depends on the care taken by the analyst in defining and measuring concepts and in interpreting conclusions drawn from the data. The authors will not persuade careful readers that they have seriously confronted important empirical and interpretive problems surrounding the measurement of the relative importance of background and environment in explaining outcomes.

First consider the definitions of the two key variables used in Herrnstein and Murray's empirical study: IQ and family background. In their empirical analysis, the operational definition of IQ is the AFQT score. Their treatment of family background is cavalier, to put it mildly. However, it is not much more cavalier than many papers on status attainment in sociology. The index is based on parental education and occupational status, and on family income measured at one point in the life cycle of the child. For many persons, the family income measure is entirely missing and is omitted from the construction of the index. No sensitivity analysis is presented to allay concerns about the sensitivity of their estimates to the application of their unusual imputation procedure.

Recall that the IQ test is administered to youths aged 15–23 years. The IQ measured after early childhood may reflect the outcome of social and cultural influences. The authors attempt to eliminate these influences by using a logistic regression analysis.

It would be miraculous if 15–23 years of environmental influences could be summarized by a composite of education, occupation, and family income measured in one year. If environment is poorly measured, Herrnstein and Murray's evidence that IQ has a stronger impact on socioeconomic outcomes than their measure of environment could rise for spurious reasons. We have already seen that education affects test scores. With a standard errors-in-variables argument, their measure of IQ may proxy the mismeasured environmental variable, and if so, the importance of IQ will be overstated. This argument is especially relevant for their analysis of schooling decisions since Neal and Johnson establish that schooling raises their measure of IQ.

Similar remarks apply to the authors' study of racial and ethnic differentials in socioeconomic outcomes. If racial differentials in environments affect ability and influence measured test scores, as Moore (1986) and others have found, evidence that racial differentials weaken when ability is accounted for using regression methods does not rule out an important role for the environment in explaining performance in society. In the presence of measurement error in the

environmental variables, the authors' analysis will overstate the "true" effect of ability on those outcomes.[10]

Another controversial feature of the authors' analysis is their method for judging the importance of the background and ability variables in a logistic regression. They do not use conventional but controversial methods such as "percentage of variance explained" or "relative size of slope coefficients" to compare the relative importance of two measures. The former method is well known to give nonunique decompositions due to nonorthogonality of regressors. The latter is known to be dependent on units of measurement (see, e.g., Goldberger 1968).

The authors construct the following unit-free measure. For two variables X_1, X_2, fix X_1 at its median and vary X_2 over a 95 percent range, from the top 2.5 percentile to the bottom 2.5 percentile. Then fix X_2 at its median and vary X_1 over the same range. This leads to two differences:

$$P(X_1^{\text{median}}, X_2^{2.5}) - P(X_1^{\text{median}}, X_2^{97.5}), P(X_1^{2.5}, X_2^{\text{median}}) - P(X_1^{97.5}, X_2^{\text{median}}),$$

where P is the logit function, and X_k^j denotes the jth percentile value of X_k. The variable with the greater range is said to have the greater explanatory power. Unfortunately, this comparison depends on the level at which the variable not being changed is set.

A cleaner measure would have been based on the logits of P ($\ln[P/(1 - P)]$), which, under their specification, are linear in X_1 and X_2. Corresponding variations in these logits are not dependent on the point of evaluation of the variable not being changed. Variations in X_1 and X_2 over 95 percent of their range are well defined and unit-free. With this transformation, the estimates reported by Herrnstein and Murray support the claim that a 95 percent variation in background produces a smaller change in the logits of P than a 95 percent variation in IQ. This is true for *all* variations in IQ and family background over any range.[11]

Two considerations limit the usefulness of the authors' method of

[10] Cavallo et al. (1995) demonstrate an important error in the Herrnstein and Murray analysis of racial wage differentials. In the analysis presented in app. 5, they pool men and women and claim to eliminate racial wage differentials by introducing AFQT into log wage equations. When dummy variables for gender are introduced into their wage equation, a racial wage differential reappears, albeit in attenuated form. The analysis of Neal and Johnson (1994) also finds important racial differentials in wages after controlling for AFQT.

[11] Herrnstein and Murray confuse matters by claiming that they are using two-standard-deviation ranges in X_1 and X_2 and that the point of evaluation of the variable being held constant is the mean. Actually both variables used in their analysis are normally distributed, and hence the mean equals the median and a two-standard-deviation spread corresponds to an approximately 95.4 percent change from the top 2.5 percentile to the bottom 2.5 percentile (see Cawley et al. 1995).

measuring the relative importance of variables in a regression. First X_1 and X_2 may not be percentile comparable. Suppose, for example, that both are dichotomous variables with different percentages in the one and zero statuses. Then for most percentiles, their comparison would not be defined, and it is very likely that there are no percentiles for which it is defined. Second, the appropriate economic measure of importance of two variables is the relevant marginal cost for achieving a given change in the outcome. The variable X_1 may be more important in their sense than the variable X_2, but the latter may be the more cost-effective intervention. The authors do not provide any justification for their measure of relative importance. Any such measure that is divorced from cost considerations is hard to interpret. In a Cobb-Douglas production function, is labor or capital more important? Both are required to produce output, and the effect of one input depends on the level of the other. In equilibrium, their relative productivity is the same as their relative marginal factor costs.

E. The Disjointed Nature of the Argument

Finally, the book fails because it does not cumulate in a convincing way. One could make a powerful case against affirmative action and in favor of testing in the workplace without appealing to g or its genetic determinants. Once the superficiality of Herrnstein and Murray's cost-benefit analysis of social programs is recognized, their long digression on genetics is seen to be irrelevant to any serious assessment of social policies.

The argument in the book does not support either their science fiction scenario of total cognitive stratification or the alternative scenario of communitarian salvation that concludes the book. No evidence is presented that sorting and heritability mechanisms operate so strongly that rigid cognitive classes have formed or are likely to form. The authors ignore their own evidence of much socially productive behavior even among low-ability groups. They implicitly assume but nowhere justify the claim that technical progress will inexorably render low-ability persons economically obsolete. They underestimate the capabilities of entrepreneurs and markets to adjust to technical change in the long run and to effectively utilize unskilled labor. The record of twentieth-century economic history does not support the view that skill differentials must inevitably increase or that technologies cannot be developed for efficiently utilizing low-skill labor.

Herrnstein and Murray's hope for a communitarian solution to the problems of inequality and cognitive partitioning is even less well grounded. They do not discuss mechanisms or incentives that would

lead persons to voluntarily sort themselves back into cognitively integrated local neighborhoods. Nowhere do they document that cognitive integration of communities is Pareto optimal. It may benefit the less able, but there is no apparent gain for the more able. The evidence on peer effects in schools is inconclusive at best (Heckman and Neal 1995).

A major theme of early twentieth-century American literature was the stifling nature of small-town life and the benefits and opportunities of the city, where persons could sort with like-minded souls. Herrnstein and Murray do not explore the social and economic forces that give rise to social partitioning, such as it is. Such sorting may have arisen from spontaneous forces of self-selection and voluntary choice. Imposing cognitive integration may require coercion from governments, a position that a libertarian such as Murray would be unlikely to espouse in other contexts.

Other policies advocated by the authors do not require acceptance of *g*, the importance of genetics, or the desirability of movement toward a cognitively integrated society. Effective arguments for simplifying laws and regulations, eliminating barriers to competition, and encouraging responsible social behavior can be and have been made without resorting to genetics or psychometrics.

III. Lessons and Challenges for Research

In spite of the disjointed nature of its argument and the careless way data and published evidence are handled, there are important lessons to be learned from the *Bell Curve*. Psychometrics is not a fraud. Tests predict productivity in schools and in the market, and they predict performance in society at large. One linear combination of test scores, *g*, does remarkably well in synthesizing what a battery of tests predict. Part of the disparity in the performance of demographic groups in schools and the workplace is due to differences in ability that can be measured by tests taken at age 17.

Nonetheless, if one uses the statistical standards advocated by the authors, more than one combination of tests is statistically significant in explaining wages. Moreover, much of the variance in social performance is not predicted by the entire set of available test scores. Many other aspects of individual skills and motivation explain performance in society at large.

Nothing in this book or in the published literature resolves the nature-nurture debate. The authors' measure of IQ comes too late in life to be informative on this issue. Their measure of environmental influence is too crude to provide a credible counterpart to their measure of IQ. A convincing analysis of the determinants of test scores

at age 17 is not presented in this book or in the published literature. Even if the authors had resolved the contribution of genes to cognitive ability, such information, by itself, is of no value in assessing the effectiveness of social programs designed to enhance skills and improve performance in society.

Nonetheless, Herrnstein and Murray's evidence and the evidence assembled from studies of many government skill remediation programs suggest that persons from disadvantaged environments are not very malleable after their late teens or perhaps their early twenties. Successful interventions for such people are known to be costly. The available evidence on the effectiveness of training programs supports the view that such efforts make only modest changes (see, e.g., Heckman 1994; LaLonde 1995). The literature suggests a particularly poor performance of educational remediation programs for adults of low ability as measured by AFQT and other cognitive tests.

Interventions are most likely to be effective when they are applied to the young. The evidence summarized in chapter 17 of the *Bell Curve* and other evidence from high-intensity enriched environment programs points in this direction. Studies by Moore (1986) and Sowell (1994) stress the role of culture and values in shaping the expectations and motivations of young children.

Job training and education are generally wasted on low-IQ and poorly motivated adults. However, subsidies for employment may be justified for them. Investments are most efficiently made in the young. There is much evidence that learning begets learning. It is much easier and more cost effective to train a young child than an illiterate young adult. However, the successful programs are often motivational in nature. Basic IQ is difficult to budge, but much evidence from civilian and military life argues that work experience has an important role to play in determining successful performance in society. Studies of productivity in civilian life indicate that motivation and personality factors also play important roles.

The available evidence indicates that early interventions are the most effective ones. Yet this evidence is fragmentary. Future research would benefit from adopting a life cycle perspective that focuses on assessing the determinants of growth and development in ability and motivation prior to age 14, the youngest age of persons in the Herrnstein-Murray sample.

Heterogeneity in endowments and life experience among persons produces a wide range of skills and motivations. Application of race-neutral, productivity-related criteria in employment and admissions decisions will lead to disparity in outcomes among demographic groups. Use of demographic quotas in hiring and admissions decisions is likely to reduce economic efficiency. Credible estimates of the

costs of using quotas do not exist. If available, they would inform the current debate about the cost of policies that insist on equality of outcomes rather than equality of opportunity. Such evidence would move the discussion of this controversial issue from an ideological low ground to an empirical high ground.

References

Angrist, Joshua D., and Krueger, Alan B. "Does Compulsory School Attendance Affect Schooling and Earnings?" *Q.J.E.* 106 (November 1991): 979–1014.

Becker, Gary S. "Irrational Behavior and Economic Theory." *J.P.E.* 70 (February 1962): 1–13.

Bound, John; Brown, Charles; Duncan, Greg J.; and Rodgers, Willard. "Measurement Error in Cross-Sectional and Longitudinal Labor Market Surveys: Validation Study Evidence." In *Panel Data and Labor Market Studies,* edited by Joop Hartog, Geert Ridder, and Jules Theeuwes. Amsterdam: North-Holland, 1990.

Cameron, Stephen V., and Heckman, James J. "Life Cycle Schooling and Educational Selectivity: Models and Evidence." Manuscript. Chicago: Univ. Chicago, Dept. Econ., 1993. (*a*)

———. "The Nonequivalence of High School Equivalents." *J. Labor Econ.* 11, no. 1, pt. 1 (January 1993): 1–47. (*b*)

Carroll, John B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies.* Cambridge: Cambridge Univ. Press, 1993.

Cavallo, Alexander; Ballinger, T. P.; El-Abbadi, Hazeem; and Heeb, Randall. "Racial Earnings Inequalities: A Critique of the Bell Curve." Manuscript. Chicago: Univ. Chicago, Dept. Econ., March 1995.

Cavelli-Sforza, Luigi L., and Bodmer, Walter F. *The Genetics of Human Populations.* San Francisco: Freeman, 1971.

Cawley, John; Heckman, James J.; Hindanov, Dmitry; and Vytlacil, Edward. "Measuring the Effects of Cognitive Ability." Manuscript. Chicago: Univ. Chicago, Dept. Econ., 1995.

Goldberger, Arthur S. *Topics in Regression Analysis.* London: Macmillan, 1968.

Goldin, Claudia, and Margo, Robert A. "The Great Compression: The Wage Structure in the United States at Mid-century." *Q.J.E.* 107 (February 1992): 1–34.

Gould, Stephen J. *The Mismeasure of Man.* New York: Norton, 1981.

Grusky, David N., and DiPrete, Thomas A. "Recent Trends in the Process of Stratification." *Demography* 27 (November 1990): 613–37.

Harris, Dickie; McCloy, Rodney; and Statman, Mary Ann. "Abilities, Job Experience and Job Performance." In *A Conceptual Framework for Analyzing and Developing a Compensation System,* edited by Paul Hogan and Timothy Dall. Fairfax, Va.: Lewin-VHI, 1995.

Hartigan, John A., and Wigdor, Alexandra R., eds. *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery.* Washington: Nat. Acad. Press, 1989.

Heckman, James J. "Is Job Training Oversold?" *Public Interest,* no. 115 (Spring 1994), pp. 91–115.

———. "Cracked Bell." *Reason* 27 (March 1995): 49–56.

Heckman, James J., and Neal, Derek. "James Coleman's Contribution to

Education: An Intellectual Odyssey." In *James S. Coleman: Falmer Sociology Series,* edited by Jon Clark. London: Falmer, 1995.
Heckman, James J.; Roselius, Rebecca; and Smith, Jeffrey. "U.S. Education and Training Policy: A Re-evaluation of the Underlying Assumptions behind the New Consensus." In *Labor Markets, Employment Policy, and Job Creation,* edited by Lewis C. Solmon and Alec R. Levenson. Boulder, Colo.: Westview, 1994.
Herrnstein, Richard J. *IQ in the Meritocracy.* Boston: Little Brown, 1973.
Herrnstein, Richard J., and Murray, Charles. *The Bell Curve: Intelligence and Class Structure in American Life.* New York: Free Press, 1994.
Herrnstein, Richard J.; Nickerson, Raymond S.; de Sánchez, Margarita; and Swets, John A. "Teaching Thinking Skills." *American Psychologist* 41 (November 1986): 1279–89.
Holland, Paul W., and Rosenbaum, Paul R. "Conditional Association and Unidimensionality in Monotone Latent Variable Models." *Ann. Statis.* 14 (December 1986): 1523–43.
Hunt, Earle. "Mokita and the Living Dog: Comments on the Role of Intelligence in Modern Society." Manuscript. Seattle: Washington Univ., Dept. Psychology, 1994.
Hunter, John E., and Schmidt, Frank L. "Fitting People to Jobs: The Impact of Personnel Selection on National Productivity." In *Human Performance and Productivity,* vol. 1, *Human Capability Assessment,* edited by Marvin D. Dunnette and Edwin A. Fleishman. Hillsdale, N.J.: Erlbaum Assoc., 1982.
Jacoby, Russell. *The Last Intellectuals: American Culture in the Age of Academe.* New York: Basic Books, 1987.
Jensen, Arthur. "*g*: Artifact or Reality?" *J. Vocational Behavior* 29 (1986): 301–31.
Joynson, Robert B. *The Burt Affair.* London: Routledge, 1992.
Kamin, Leon. *The Science and Politics of IQ.* Potomac, Md.: Erlbaum Assoc., 1974.
Kaus, Mickey. *The End of Equality.* New York: Basic Books, 1992.
LaLonde, Robert. "The Promise of Public Sector–Sponsored Training Programs." *J. Econ. Perspectives* 9 (Spring 1995): 149–68.
Mare, Robert D. "Social Background and School Continuation Decisions." *J. American Statis. Assoc.* 75 (June 1980): 295–305.
———. "Five Decades of Educational Assortative Mating." *American Sociological Rev.* 56 (February 1991): 15–32.
Mincer, Jacob. *Schooling, Experience, and Earnings.* New York: Columbia Univ. Press (for NBER), 1972.
Moore, Elsie G. J. "Family Socialization and the IQ Test Performance of Traditionally and Transracially Adopted Black Children." *Developmental Psychology* 22 (May 1986): 317–26.
Murnane, Richard; Levy, Frank; and Willett, John. "The Growing Importance of Cognitive Skills in Wage Determination." *Rev. Econ. and Statis.* (Spring 1995).
Murray, Charles A. *Losing Ground: American Social Policy, 1950–1980.* New York: Basic Books, 1984.
———. *In Pursuit: Of Happiness and Good Government.* New York: Simon and Schuster, 1988.
Neal, Derek, and Johnson, William. "The Role of Pre-market Factors in Black-White Wage Differences." Manuscript. Chicago: Univ. Chicago, November 1994.
Olea, Michele Morales, and Ree, Malcolm James. "Predicting Pilot and Navi-

gator Criteria: Not Much More than *g*." *J. Appl. Psychology* 79 (December 1994): 845–51.

Phelps, Edmund S. "A Program of Low-Wage Employment Subsidies." Manuscript. New York: Sage Found., 1994.

Ree, Malcolm James, and Earles, James A. "Intelligence Is the Best Predictor of Job Performance." *Current Directions Psychological Sci.* 1 (June 1992): 86–89.

Ree, Malcolm James; Earles, James A.; and Teachout, Mark S. "Predicting Job Performance: Not Much More than *g*." *J. Appl. Psychology* 79 (August 1994): 518–24.

Reich, Robert B. *The Work of Nations: Preparing Ourselves for 21st-Century Capitalism.* New York: Knopf, 1991.

Schmidt, Frank. "The Problem of Group Differences in Ability Test Scores in Employment Selection." *J. Vocational Behavior* 33 (1988): 272–92.

Schweinhart, Lawrence J.; Barnes, Hazel; and Weikart, David P. *Significant Benefits: The High-Scope Perry Preschool Study through Age 27.* Ypsilanti, Mich.: High-Scope Press, 1993.

Sowell, Thomas. *Race and Culture: A World View.* New York: Basic Books, 1994.

Spearman, Charles S. "'General Intelligence,' Objectively Determined and Measured." *American J. Psychology* 15 (April 1904): 201–93.

Sternberg, Robert J. "General Intellectual Ability." In *Human Abilities: An Information-Processing Approach,* edited by Robert J. Sternberg. New York: Freeman, 1985.

———. *The Triarchic Mind: A New Theory of Human Intelligence.* New York: Penguin, 1988.

Suppes, Patrick, and Zanotti, Mario. "When Are Probabilistic Explanations Possible?" *Synthese* 48 (August 1981): 191–99.

Taber, Christopher. "The Rising College Premium in the Eighties: Return to College or Return to Ability?" Manuscript. Chicago: Univ. Chicago, November 1994.

Zigler, Edward, and Muenchow, Susan. *Head Start: The Inside Story of America's Most Successful Educational Experiment.* New York: Basic Books, 1992.