

Hope, Anguish, and the Problem of Our Time: An Essay on Publication of *The Black-White Test Score Gap*

SAMUEL R. LUCAS

University of California, Berkeley

The Black-White Test Score Gap, edited by Christopher Jencks and Meredith Phillips, raises a series of important questions as to the cause and consequences of observed differences in measured achievement. Jencks, Phillips, and colleagues provide the most successful and sustained assessment of the claims of racial difference attributed to Herrnstein and Murray's *The Bell Curve*. Even so, striking similarities between the two books reveal the inability of current analyses to truly deepen our understanding of race in America.

In the metropolis of the modern world, in this the closing year of the nineteenth century, there has been assembled a congress of men and women of African blood, to deliberate solemnly upon the present situation and outlook of the darker races of mankind. The problem of the twentieth century is the problem of the color-line, the question as to how far differences of race—which show themselves chiefly in the color of the skin and texture of the hair—will hereafter be made the basis to denying over half the world the right of sharing to their utmost ability the opportunities and privileges of modern civilization.

—W. E. B. Du Bois, 1900, Address to the Nations of the World

Although Du Bois identified the problem of the twentieth century with international sweep, as the century comes to a close American social scientists are still trying to determine just what the American problem is. Even a boundless optimist must be disheartened that now, nearly a century after Du Bois' call, many leading scholars in the most powerful nation the world has ever seen are still engaged in argument around the facts of racial inequality. Further, this continuing debate has created neither a new consensus nor a redoubled commitment to eradicating racial inequality; instead, in the last years of the twentieth century American scholars have resurrected an old query concerning race. Scholars now ask the tired question yet again: Is racial inequality in the United States driven by social factors, or does it flow from biogenetic differences between blacks and whites?

It is within this context, established by both 100 years of anticipation and a recent reinvasion of doubt sowed by fallout from *The Bell Curve*, that the Jencks and Phillips volume arrives. Consequently, it bears both the promise and the pathos of that history. First, the promise.

Jencks and Phillips have drawn together a strong cast of researchers to cover a vast area. They ask a series of clearly important questions. What are the gaps in test scores between blacks and whites? and Have the test score gaps widened or narrowed over the last twenty years? How have the gaps appeared to change during early childhood through adolescence? How likely is it that child-rearing practices, oppositional cultural ethos, genetic endowments, teacher expectations, teacher-student interactions, and more account for the emergence, existence, and maintenance of the black-white test score gap? How useful are test scores for allocating students to scarce slots in elite institutions? How can elite colleges and universities maintain racial diversity as affirmative action in the case of race is delegitimated (while affirmative action for athletes, faculty children, alumni children, musicians, and others remains intact and virtually uncontested)? How much labor market inequality can be explained by test score differences, if any? And they ask, How can one assess whether tests are biased against blacks? and How difficult is it to assess bias in any situation?

These are all good questions, and the analysts addressing them are for the most part successful. For example, in an important essay Richard E. Nisbett reviews the evidence for and against a genetic basis for test score differences. Nisbett shows that analysts have used several different ways to measure race, and that the majority of analyses fail to show genetically linked racial differences in test scores. This is an important paper, for most discussants on race and inequality seem to have no knowledge of the literature Nisbett reviews.

In a provocative chapter on the effects of schools in narrowing the black-white gap, Ronald F. Ferguson reviews evidence on, among other factors, teacher certification. Ferguson reasons that because black would-be teachers are less likely to pass certification exams than are white would-be teachers, certification testing should raise the quality of black teachers. And because in many jurisdictions black students are more likely to have black teachers than are white students, black students should benefit from the introduction of teacher certification testing. Ferguson presents some evidence that the adoption of teacher certification testing is associated with declines in the black-white test score gap. This kind of analysis is extremely important, for it suggests that issues surrounding equity and opportunity for different constituencies are quite complex.

Another highlight of the book is the chapter by Phillips, Brooks-Gunn, Duncan, Klebanov, and Crane. This chapter addresses the role of family background and parenting practices on the black-white test score gap. The

analysis is noteworthy because of its appropriately broad view of social background. Phillips et al. not only supplement the standard indicators of background with measures of family wealth, but also extend the analysis of background in two oft-neglected ways. First, they bring into their analysis the way in which mother's background is produced. Thus they include as predictors of a child's test score not only the mother's test scores, but also characteristics of the mother's high school that may have played a role in the production of both the test score and other important characteristics. Second, they include information on grandparents' characteristics, further elaborating the direct and indirect social background of the child. With this analysis Phillips et al. show that their expanded understanding of social background accounts for up to two-thirds of the black-white test score gap. They correctly conclude that although their analysis cannot identify the effect of environment on the black-white test score gap, it does suggest that closing the gap in social background between blacks and whites could greatly reduce the test-score divergence.

The Phillips et al. analysis is consistent with findings Grissmer, Flanagan, and Williamson report two chapters later. Grissmer et al. find that arraying test scores into cohorts (a difficult enterprise given existing data) shows that rising black achievement tracks suggestively with policy changes that have improved family environments.

Although in many ways family environments have changed for the better over the last quarter century, some commentators argue that peer-produced environments have worsened, particularly for blacks. However, in this volume Cook and Ludwig argue against the thesis that black students refuse to achieve because black students do not want their peers to charge them with "acting white." Cook and Ludwig use National Educational Longitudinal Study of 1988 (NELS) data in their analysis, which joins a growing set of research efforts that have failed to find evidence of the acting-white thesis using national data.

In an interesting comment on this chapter Ferguson maintains that the NELS data do not measure the key indicators implied by the acting-white thesis. Ferguson argues that NELS has no measure of whether a student hesitates to raise his or her hand in class or participate in class discussions. Reluctance to participate, he argues, would be a sign of effective peer pressure. Ferguson makes a plausible claim to which a vast set of analytic problems attend. The resolution of those problems will probably entail using a mix of research methods.

Claude M. Steele and Joshua Aronson propose an alternative social psychological theory to explain test performance. Focusing on the actual test administration, they investigate whether fears of confirming negative stereotypes are associated with lower test performance. Their evidence suggests that the testing situation and thus performance are greatly affected by

attitudes prevalent in the wider society. Indeed, the student more committed to success is more likely to have his or her performance affected by societal attitudes than are others. This paper raises questions concerning the future performance of students who reside in states undergoing intense efforts to reverse race-based affirmative action. If Steele and Aronson are correct, then we may expect an increase in the black-white gap in such states, especially where leading political spokespersons publicly assert that black collegians admitted prior to the end of affirmative action were unqualified.

The Black-White Test Score Gap addresses many other questions of great importance, and provides several solid answers. We learn the black-white test score gap has changed over the last twenty-five years, and that most of the change is in the direction of declining black-white differences. Despite some closure in the chasm between the groups, however, a substantial canyon remains. We learn that child-rearing practices and teacher expectations appear to be important parts of the process through which test score differences are created. We learn that even though blacks are more likely to be poor than are whites, efforts to substitute wealth or income affirmative action for race-focused affirmative action will fail to secure racial diversity on campus because the vast majority of the poor are not black. We learn that test scores seem to predict labor market outcomes such as employment and earnings, and part of the black-white gap in these important quality-of-life indices is connected to test score differences.

Because *The Black-White Test Score Gap* arrives in a context still reverberating from the clash of sentiments precipitated by the public understanding of *The Bell Curve*, a few comments on Herrnstein and Murray's research are in order. In this connection it must be noted that *The Black-White Test Score Gap* effectively addresses the issues it treats, and in doing so Jencks, Phillips, and colleagues add an important contribution to the effort that is well on the way to thoroughly refuting the claims of *The Bell Curve*. Indeed, *The Black-White Test Score Gap* provides the most successful direct and sustained attack to date on the alleged race-related claims of *The Bell Curve*, and that these authors do so on Herrnstein and Murray's chosen terrain is persuasive. I write "alleged race-related claims of *The Bell Curve*" because misreadings of the Herrnstein and Murray treatise, encouraged by media focus on a decidedly small section of the book, suggest that *The Bell Curve* argues that black-white test score differences flow from black intellectual inferiority rooted in biology. Herrnstein and Murray encourage this misreading in part by suggesting that other analysts have been too afraid to talk honestly about race, as if they alone are brave enough to state the conclusions to which scientific inquiry must lead. But Herrnstein and Murray never directly claim racial differences in test scores are driven by genetic differences.

Both the insinuated and the explicit claims—that racial differences in test scores are genetic, and that Herrnstein and Murray are the two brave

social scientists willing to talk honestly about race—would be comical were they not deeply tragic. Herrnstein and Murray use the main tool at their disposal, statistical analysis, poorly, and they do so using data constructed owing to widespread consensus that test scores are important. Were Herrnstein and Murray the only social scientists willing to consider the importance of intelligence, they would have no data with which to attempt to do the job. Further, on the point of courage, the sole surviving author refuses to discuss in public the pivotal part of *The Bell Curve*—the statistical analysis—with scholars such as Christopher Jencks, Meredith Phillips, Sanders Korenman, Christopher Winship, Robert M. Hauser, Michael Hout, Richard Arum, Charles Manski, James Heckman, Thomas Kane, Arthur Goldberger, Robert D. Mare, V. Joseph Hotz, and on and on or, in short, with anyone who routinely uses statistical methods and thus might venture to raise decisive criticisms of *The Bell Curve* analyses.

Given this context it is important to affirm that in *The Black-White Test Score Gap*, Jencks, Phillips, and their compatriots conduct much more effective statistical analyses, reference important ethnographic investigations bearing on the issue, and provide nuanced and compelling theoretical evidence as well. Thus their work supplies another important corrective for the public record.

However, although Jencks, Phillips, and their colleagues have made an important contribution that can only further our understanding of the complexity of the production of racial differences in test scores, there is a striking similarity between major portions of their work and that of *The Bell Curve*, and it is this similarity that calls forth not only hope, but also anguish, as the concluding notes of the book and of the century slowly fade away. The faith undergirding both *The Bell Curve* and *The Black-White Test Score Gap* is that if one just controls for enough of the behaviors and conditions that seem to differentiate whites and blacks, one will reach a true answer as to whether or not the racial difference in test scores is based in the genetic inferiority of blacks. For Herrnstein and Murray, the list of factors to control is short and the answer is yes, genetic differences are probably important. For Jencks, Phillips, and colleagues, the list of factors to control is long and potentially limitless, and in *The Black-White Test Score Gap* they suggest the answer is no, genetic differences are probably not important. Yet, in both *The Bell Curve* and *The Black-White Test Score Gap* the answers remain only provisional estimates of just how much blacks qua blacks and whites qua whites differ. Again, Jencks, Phillips, and colleagues have been very creative at devising and introducing plausible additional factors that may reduce black-white differences, and unlike Herrnstein and Murray they are both properly cautious in their claims and successful in connecting their claims to appropriate analyses appropriately conducted. Yet the faith appears to remain, the faith that one can either eradicate the

performance differences between blacks and whites through statistical approaches and thereby explain the gap between the groups' performance, or, if that fails, reveal an essential difference between blacks and whites that cannot be eradicated.

This faith, and this effort, is not new. Consider that in 1785 in *Notes on the State of Virginia* Thomas Jefferson made a case against black intelligence that is logically indistinguishable from the case made in the last years of the twentieth century. Jefferson argued that he had never seen blacks reason, and they seemed to have no capacity for abstract thought. Jefferson reached this conclusion by comparing black slaves and free whites, arguing that:

It would be unfair to follow them to Africa for this investigation. We will consider them here, on the same stage with the whites, and where the facts are not apocryphal on which a judgment is to be formed. It will be right to make great allowances for the difference of condition, of education, of conversation, of the sphere in which they move. Many millions of them have been brought to, and born in America. Most of them indeed have been confined to tillage, to their own homes, and their own society: yet many have been so situated, that they might have availed themselves of the conversation of their masters; many have been brought up to the handicraft arts, and from that circumstance have always been associated with the whites. Some have been liberally educated, and all have lived in countries where the arts and sciences are cultivated to a considerable degree, and have had before their eyes samples of the best works from abroad. (pp. 139–140)

Jefferson suggests that slaves and free whites are distinguished by relatively slight differences in their conditions of work and of life. Thus it appears an easy matter to mentally adjust for these differences and thereby construct a correct comparative assessment of their capabilities. Jefferson thus implies that black slaves and free whites actually occupy the same social stage by virtue of living in the same geographical place. Although the hindsight of over 200 years should show Jefferson's methods to be flawed, note the logical similarity between Jefferson's effort and current analytical approaches. The faith many analysts place in statistical methods for adjusting for differences between blacks and whites is analogous to Jefferson's attempt "to make great allowances for the difference of condition, of education, of conversation, of the sphere in which they move," as if one can really and truly equate groups with such vastly different experiences. The only dissimilarity between Jefferson's effort and that of many turn-to-the-twenty-first century social scientists is in the formalization of the adjustment; by their actions both seem to suggest such adjustments can be done effectively.

However, introductory statistical texts suggest this is not possible. Neter, Wasserman, and Kutner (1989) note “we should caution again about making estimates or predictions outside the scope of the model. The danger, of course, is that the model may not be appropriate when extended outside the region of the observations. In multiple regression, it is particularly easy to lose track of this region since the levels of X_1, \dots, X_{p-1} jointly define the region. Thus, one cannot merely look at the ranges of each independent variable” (p. 262, emphasis in original).

Neter, Wasserman, and Kutner imply that if the conditions under which black and white students experience school are qualitatively dissimilar, introducing terms into a statistical model is unlikely to make them similar enough to make the kinds of comparisons we often seek to make. We commonly introduce terms such as family income into statistical models. For this type of variable statistical control can be effective because the black-white difference in family income is a matter of degree. Blacks are on average poorer than whites, but there are some wealthy blacks and there are many poor whites. This is the kind of factor for which statistical control is useful.

However, there are many other variables that are *not* matters of degree. And it is these variables that define what it means to be black in America at this time. An example of one such variable concerns police behavior. Police do not stop whites for “driving while black,” but police do stop blacks, including wealthy blacks, for this offense. This is the kind of situation for which statistical control is problematic. One statistical reason for the problem is that such variables create analytic difficulties analogous to those generated when two variables are perfectly correlated. When two variables are perfectly correlated statistical analysis cannot disentangle their effects. Theoretically, however, controlling for such variables makes no sense, for they are the variables that define blacks socially in this society. Thus it would be wiser to regard “driving while black” and being black not as two variables but, instead, as part of the same condition. It is this second type of variable that forces one to conclude that by definition blacks and whites do not occupy the same social space.

There are many variables of the second type, and collectively they render the effort to statistically equate blacks and whites unsuccessful. Certainly adjustments for factors such as family income should reduce the gap between the two races under discussion, but the continued presence of definitive, qualitative differences in experience destroys the possibility of statistically equating the two groups. Hence, analyses with statistical controls will usually reduce the observed gap, but the remaining gap will likely still be larger than it would be were one able to truly equate the two groups. All one can correctly say after analyses such as those in *The Black-White Test Score Gap* is that some factors seem to mediate or account for some of the observed

difference, but one cannot infer that the remaining difference is unexplained by social factors. However, the vacuum created by the remaining and seemingly ineradicable black-white gap is frequently filled by the inference that the unexplained gap exists owing to the genetic basis of race-linked performance differences. Hence, the oft-ignored limitations of statistical analyses not only reduce the utility of statistics for ending suspicion that black genetic inferiority explains blacks' relatively lower test scores but also can easily and erroneously strengthen the argument for genetic-based differences in intelligence, because statistical analyses are often not properly contextualized and because pivotal problematic assumptions often remain undiscussed.

This state of affairs has an important implication, I think, for the work presented in *The Black-White Test Score Gap* and for continued investigation of racial inequality. It suggests that analysts need to look beyond the formal thought experiments embodied in statistical models. To understand how race remained the problem of the twentieth century in America throughout the twentieth century one needs to investigate how it is that race comes to be implicated in allocating resources to different groups. One needs to investigate what taken-for-granted practices sediment racial division in the very process of resource allocation. This effort requires an initial move beyond statistics. A illustrative case will suggest what I mean, and also point toward a vast, promising, and relatively unexplored area for analytic work around testing.

I take the concept of sedimentation from Oliver and Shapiro (1995), whose analysis of black and white wealth provides a model for researchers interested in incorporating and moving beyond statistics in analyzing racial differences. For the question of test score differences some of the important groundwork for such an approach is laid very effectively in *The Black-White Test Score Gap*. In a chapter titled "Racial Bias in Testing" Jencks provides a helpful typology of the kinds of biases tests might and might not have, and the kinds of evidence one would need to consider to support a contention of racial bias. Further, he makes a case that blacks are disadvantaged because the technology for constructing tests of cognitive achievement is well developed and thus tests have less error than do the measuring devices for factors such as motivation, persistence, and such. Thus, Jencks argues, when we rely on tests to allocate persons to educational and occupational positions because tests have less error than other measures, we make blacks pay for the greater investment we have made in constructing cognitive tests.

But is our current testing technology actually advanced, or is it really constructed on a series of analytically problematic procedures whose primary social effect is to legitimate the use of tests, the existing allocation of power, and the transference of that power to the children of the powerful?

I suspect it is more the latter than the former, and a brief description of an important aspect of the process of test construction will make the case. In essence, key procedures in test construction operate like the house odds in Las Vegas, subtly but powerfully stacking the deck against disadvantaged groups as such while allowing some members of those same groups to experience windfall gains. Indeed, enough members of disadvantaged groups may experience enough windfall gains to keep everyone coming back faithfully year after year to play the same game again and again.

The illustration I want to make concerns an important principle of test design. In order to construct a standardized test item writers draft a set of candidate questions and administer them to a test-taking population. For the SAT-I the administration of candidate questions is typically done as part of the testing process, such that every SAT-I test-taker answers some candidate questions that will be evaluated for future use. Test-takers' performance on candidate questions are not used in the calculation of their scores.

After the testing has been completed, analysts evaluate how the candidate questions performed. A key aspect of the evaluation considers which students answered the candidate questions correctly. If test-takers who obtained low scores on the existing test were more likely to answer a candidate question correctly than did test-takers who obtained high scores on the existing test, then the candidate question is rejected because it is seen as failing to differentiate appropriately between high and low scorers. In other words, if the "best and brightest" (as determined by performance on the existing test) cannot answer the candidate question, but the "dullards" can, then the question is rejected by test-makers. That test-makers use this principle to reject candidate questions every year suggests that test construction procedures may very well penalize many students from historically poorly performing groups unfairly. Whether or not this procedure *does* penalize students from poorly performing groups is an empirical question. But it is worth noting that it is a curious task that becomes defined as evidence of knowledge on the basis of who performs the task successfully, and it is worth considering what this means for the entire process of testing.

This procedure is problematic for many policy questions, but with specific reference to the black-white test score gap the implications are many, subtle, and potentially important. Any procedure that rejects a question that students on the bottom of the test score distribution are more likely to answer correctly than those on the top simply *because* those on the bottom of the test score distribution are more likely to answer the question correctly than those on the top is, by definition, discriminatory. The procedure is discriminatory because it celebrates or disregards achievement simply by virtue of who accomplished the achievement. That this procedure is accepted virtually without reflection by many test-makers suggests just how

discrimination against the disadvantaged is sedimented into the standard operating procedures of many of our institutions.

Note also that this procedure is not necessarily *racially* discriminatory. However, given that on prior tests black students have scored lower on average than have white students, the procedure of not counting a question when students on the bottom of the previous tests' distribution outperform students on the top of the previous tests' distribution may have a disparate and negative impact on black students' scores.

If the principle of rejecting questions on which "dullards" outperform others is obviously problematic, it must have *some* redeeming feature that sustains its use. It does. The principle is useful because analysts use test score stability and aggregate distributional stability to determine whether the test construction effort has succeeded, and they point to test score stability and aggregate distributional stability when they want to show important constituencies that the test is "working." The ability to reject questions on which "dullards" outperform others is important to securing the much-desired distributional stability.

However, the appeal to test score stability is not innocuous; it is instead a socially consequential strategy, for it means that each new version of the test must produce aggregate patterns very similar to those the previous versions provided. This means that the "best and brightest" will likely look very similar socio-demographically from year-to-year. Indeed, if after the test is administered the resulting score distributions actually *do* differ appreciably from prior distributions, then suspicions of cheating and adjustments in scores are the likely response of test-makers. If one doubts this implication, consider the experience of the students of Jaime Escalante. These students from the barrio worked hard enough to attain high scores on an Advanced Placement Calculus exam. The response of the testing company was to suspect them of cheating and demand they take the test again under the watchful eyes of test company personnel.

The aim for distributional stability, and the principles of test construction used to create distributional stability, make testing work to ease the intergenerational transfer of power. Indeed, this principle of test construction not only eases intergenerational transfers, but also produces tests that mask changes that might be occurring in the true distribution of achievement in the population. In this way this principle is a key part of the process that legitimates the existing distribution of power.

It is worth noting that there are test construction procedures less prone to this problem, and the tests used as part of the National Assessment of Educational Progress (NAEP) trend assessment are examples. In order to construct tests with less reliance on the discriminatory principle, item-writers define the domain of the content area. They then construct candidate items, draw on the judgments of experts, and in this manner determine

which concepts and questions are likely to be more or less difficult. Although additional highly detailed work is needed to better determine the difficulty of items and the cognitive processes test-takers use (e.g., Hamilton 1997), the above approaches are a step in the right direction. Moreover, the example of such tests shows that it *is* possible to greatly reduce the use of what should be an obviously discriminatory procedure for test construction, namely, the use of an item's ability to preserve the prior distribution of test-takers in evaluations of item validity.

Unfortunately, although they are partially determinative of the data that are the focus of *The Black-White Test Score Gap*, test construction procedures remain a relatively unanalyzed phenomenon. Certainly the issue posed by the discriminatory principle in item selection is not racial bias per se but, instead, whether assumptions inherent in some test construction strategies preordain that test results will mirror the past in the aggregate. However, if test construction strategies do preordain that test results largely mirror the past, then our understanding of the size of the black-white *achievement* gap and the pace of its change is likely to be wrong. Hence the very gaps that motivate the agenda of *The Black-White Test Score Gap* may be based at least as much in unanalyzed procedures of test construction as in any factor discussed in the over 500 pages of the admittedly insightful book.

Test construction needs to become a focus of sustained sociological inquiry before we conduct many more statistical analyses of students' scores. Indeed, our lack of sustained analyses of these procedures occasionally leads us to treat tests in problematic ways. For example, some authors in the Jencks and Phillips volume transform scores from several different studies into standard deviation units to compare them. Yet can test scores from disparate sources be made comparable at all absent analysis of the procedures used and decisions made in constructing the tests? The answer to this question is yes only if one believes the items on the tests do not really matter, either because all tests necessarily tap the same construct, or because the process of test construction makes the items irrelevant. But I believe that the items on a test *do* matter, and both the elaborate process of test construction and National Academy of Science analyses suggest that the items matter as well (e.g., Feuer et al., 1999). If the particular items are important, the process of item selection is, too. Thus until we develop a thoroughgoing understanding of just how the process of item selection can insure that previous differences become contemporary differences, and integrate that understanding into our analyses, we will be unable to construct a complete answer to the black-white test score gap. In other words, we must pay as much attention to the processes that create the left-hand dependent variable side of the equation as we have paid to developing the right-hand independent variable side of the equation.

Test construction is a technical enterprise, but it is also a site of power. Thus in conducting research on test construction we will need to consider historic and contemporary configurations of power. In order for these investigations to avoid the Scylla of assuming a malicious cabal at the heart of an ongoing effort to oppress certain groups and the Charbydis of presuming that existing inequality is so irrelevant to political possibility that persons need change nothing but their minds in order to provide true equal opportunity, this investigatory effort must delve into the many ways that standard operating procedures are constructed, justified, and maintained. It will need to link study of those mechanisms to individual and organizational incentives and also to the activities of persons in a variety of spheres. Only after we analyze these processes and interests will we be able to place the findings from *The Black-White Test Score Gap* into proper perspective.

However, for investigation of the right-hand side of the equation I can think of no better place to start than *The Black-White Test Score Gap*. I highly recommend this book to those interested in addressing this pressing issue of our time.

In closing, reading the Jencks and Phillips volume in the context of Du Bois' assignment for the twentieth century, it seems clear that century's work is not yet done. And, going back another hundred years or so, it seems that work is if anything even harder now than it was previously. For before alleging the intellectual weakness of blacks, and linking that weakness to general black inferiority engendered by biological factors, Jefferson noted that a Virginia bill to free the slaves called for colonizing blacks in a separate place and recruiting white immigrants to do the work slaves were forced to do. Jefferson noted:

It will probably be asked, Why not retain and incorporate the blacks into the state, and thus save the expence of supplying, by importation of white settlers, the vacancies they will leave? Deep rooted prejudices entertained by the whites; ten thousand recollections, by the blacks, of the injuries they have sustained; new provocations; the real distinctions which nature has made; and many other circumstances, will divide us into parties, and produce convulsions which will probably never end but in the extermination of the one or the other race. (p. 138)

If we hope to avoid Jefferson's dire prediction of national self-destruction we need speak frankly about the issues concerning race. But a frank discussion of race, or any other dimension of oppression, cannot be had absent conversation concerning power—its material basis, its effort to legitimate and thus mask itself, and its stubborn tendency to transform itself and remain in the same hands year after year. Thus the problem of our

time, with all its promise and danger, is the twentieth century problem writ large: the problem of the twenty-first century is the problem of the power divide—how will power be limited, how will power be shared, and how will power be used either to nurture the capacities of all, or, instead, stifle the potential of most? To the extent that *The Black-White Test Score Gap* maps the contours through which power will move, the impediments to emancipatory action erected in black *and* white culture, and the extent to which even on this challenging terrain some do succeed, the book is an important work and an important opportunity not to be missed. Yet it is also clear in the uncharted territory that remains that other, very different, and extremely important work is yet to be done.

I thank Kathleen Dzibur and Claude S. Fischer for comments on an earlier draft.

References

- Educational Testing Service. (1997). *NAEP 1996 Trends in Academic Progress*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- Feuer, M. S., Holland, P. W., Green, B. F., Bertenthal, M.W., & Hemphill, F. C. (1999). *Uncommon Measures: Equivalence and Linkage Among Educational Tests*. Washington, DC: National Academy Press.
- Hamilton, L. S. (1997). *Construct Validity of Constructed Response Assessments: Male and Female Science Performance*. Palo Alto, CA: Stanford University School of Education.
- Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: The Free Press.
- Jefferson, T. [1785] (1955). *Notes on the State of Virginia*. Chapel Hill: University of North Carolina Press.
- Jencks, C., & Phillips, M. (1998). *The Black-White Test Score Gap*. Washington, DC: Brookings Institution Press.
- Neter, J., Wasserman, W., & Kuttner, M. H. (1989). *Applied Linear Regression Models* (2nd ed.). Homewood, IL: Irwin.
- Oliver, M. L., & Shapiro, T. M. (1995). *Black Wealth/White Wealth: A New Perspective on Racial Inequality*. New York: Routledge.

SAMUEL R. LUCAS is Associate Professor in the Department of Sociology at the University of California—Berkeley, and an affiliate of the University of California's Survey Research Center and the Department of Demography. His research interests are social stratification, sociology of education, research methods, and research statistics.