

The Campaign for Educational Equity

TEACHERS COLLEGE COLUMBIA UNIVERSITY

Reassessing the Achievement Gap: Fully Measuring What Students Should Be Taught in School

By
Richard Rothstein*
Rebecca Jacobsen**
Tamara Wilder***

This is a summary of two reports prepared for the Campaign for Educational Equity at Teachers College, Columbia University. The full reports will be published shortly in book form. Nothing herein may be quoted without the express permission of the authors.

* Research Associate, Economic Policy Institute; ** Assistant Professor, School of Education, Michigan State University; *** Ph.D. candidate, Department of Organization and Leadership, Teachers College, Columbia University

In today's presentation, we summarize two products of the ongoing work we are doing for the Campaign for Educational Equity. First is "***A Report Card on Comprehensive Equity: Racial Gaps in the Nation's Youth Outcomes.***" A copy of the full report can be found at http://www.epi.org/content.cfm/racial_gaps. Second is a proposal on redesign of the National Assessment of Educational Progress, for the purpose of collecting data on the full range of youth outcomes to be sought by schools and other institutions of youth development. This report will be published in full in the near future.

The authors, and the Campaign for Educational Equity, acknowledge with gratitude the support for this research that has been provided by the Lumina Foundation. However, the analyses and conclusions of these reports are the responsibility of the authors alone, and do not necessarily represent formal positions of the Campaign for Educational Equity, Teachers College, Columbia University, the Lumina Foundation, or other institutions with which the authors are affiliated.

The authors also acknowledge with gratitude the consistent support and encouragement we have received for this project from the leadership of Teachers College. In particular, we are grateful to Arthur Levine and to Susan Fuhrman, the former president and president of Teachers College, and to Michael Rebell, the director of the Campaign for Educational Equity.

Part I:
Summary of
A Report Card on Comprehensive Equity: Racial Gaps in the Nation's Youth Outcomes.

Contemporary school accountability policy, including *No Child Left Behind* (NCLB) and similar state systems, holds all elementary schools, regardless of student characteristics, accountable for achieving proficient student scores in reading and math. By demanding that schools report achievement for racial, ethnic, and economic subgroups, the accountability system aims to shine a light on schools that "leave children behind."

At first glance, this approach seems reasonable. But recognition is growing that this accountability system has begun to shift how we think about what schools should do. By basing sanctions solely on math and reading scores, the law creates incentives to limit -- or in some cases to eliminate entirely -- time spent on other important objectives of schooling. This reorientation of instruction disproportionately affects low-income and minority children, so achievement gaps may actually widen in areas for which schools are not now being held accountable.

Last year, the Center for Education Policy (CEP) surveyed 349 representative school districts. It found that, since the enactment of NCLB, 62 percent of these districts had increased the amount of time devoted to English and math. The increases were greatest in urban districts and especially in districts which were subject to sanctions under NCLB because their test scores were too low.

This is a good thing, and is just what NCLB was intended to accomplish. Students whose reading and math performance was lowest were getting more instruction in these

subjects because their schools were threatened with sanctions if scores didn't improve. Noting the law's apparently salutary effects, U.S. Secretary of Education Margaret Spellings boasted "I'm a what-gets-measured-gets-done kind of gal."

But what doesn't get measured doesn't get done. The CEP survey found that almost all districts which had at least one school subject to NCLB sanctions because of low scores had increased instructional time in reading, by an average of three hours a week. They increased instructional time in math by an average of almost an hour and a half.

But the time had to come from somewhere. In these districts, social studies and science instruction was cut by an hour and a half each, art and music and physical education by an hour each, and recess by another hour.

As one Teachers College student, formerly a teacher in a school serving low-income minority students in Los Angeles, put it:

[T]he pressure became so intense that we had to show how every single lesson we taught connected to a standard that was going to be tested. This meant that art, music and even science and social studies were not a priority and were hardly ever taught. We were forced to spend ninety percent of the instructional time on reading and math. This made teaching boring for me and was a huge part of why I decided to leave the profession.

The Report Card on Comprehensive Equity attempts to address this distortion of the goals of education by describing the nation's gap in performance between white and black youth on a broad set of measures, not standardized test scores alone. Discussion of the achievement gap is commonplace in American policy making today, but typically, the gap is defined as the difference in standardized test scores, usually only in math and reading, between black and white students.

Because ‘what gets measured gets done,’ the Report Card on Comprehensive Equity proposes to measure the neglected outcomes of public education.

This report differs from typical discussions of the achievement gap in these ways:

* The Report Card describes eight broad goals of American schools and other institutions of youth development, and suggests that the ideal of comprehensive equity must include a narrowing of the performance gap between black and white youth in each of these eight goal areas. These goals have historically been central to Americans' conception of education and youth development and remain so today.* They are a meaningful opportunity to achieve adequate

basic academic skills,

critical thinking and problem solving,

social skills and work ethic,

readiness for citizenship and community responsibility,

foundation for lifelong physical health,

foundation for lifelong emotional health,

appreciation of the arts and literature, and

preparation for skilled work, for those youths not destined for academic college.

* The Report Card focuses, to the extent possible, on the *outcomes* of educational and youth development institutions. "Outcomes" are the achievement of adequacy in

* For a brief review of the historical development of education goals, see Richard Rothstein, and Rebecca Jacobsen. 2006. "The Goals of Education." *Phi Delta Kappan* 88 (4), December.

these eight goals by adolescents and young adults, roughly between the ages of 17 and 25, who are at an age to have successfully completed the normal course of elementary and secondary education and to exhibit, in their behaviors, the impact of our educational and youth development policies.

Typical discussions of the performance gap focus on test scores during students' school careers. Such scores are only predictors, to an unknown extent, of academic achievement at the end of the schooling years, which is the true object of education and youth development. The Report Card terms "outputs" these scores, as well as other markers of youths' accomplishments during the school years. It distinguishes such outputs from "outcomes," which describe performance at the conclusion of the educational and youth development process. The Report Card includes measurement of outputs only to supplement data on outcomes, because available outcome data are inadequate to tell a complete or fully accurate story.

Because both outcome and output data on each of the broad goal areas, while suggestive, are not definitive, even at the national level, the Report Card on Comprehensive Equity is also based on estimates of black-white gaps in a collection of inputs that research and common sense suggest have a positive impact on performance in each of the goal areas. These inputs, again limited by the availability of representative national data, include parental involvement in their children's schooling; parental engagement in challenging activities with their children; the avoidance by children of excessive television watching; young children's attendance at high-quality preschools; students' exposure to positive peer influences; children's access to books both in public libraries and at home; and a collection of fiscal resources including school spending,

smaller classes for young children, better-paid teachers, relatively strong state fiscal support for higher education.

Although the Report Card makes no claim to complete precision in any of these estimates, our judgment is that a measurement of gaps in each of the eight goal areas should incorporate this gap in inputs likely to produce more adequate outcomes. Therefore, the Report Card's estimates of the black-white gap in each of the eight goal areas include, with a weight of 20 percent in each case, this overall input gap. The performance gap described in each of the eight goal areas is comprised of 80 percent of gaps in the various outcome and output indicators of performance in that goal area, and 20 percent of the gap in overall inputs.

* Many indicators of black-white differences, used in the Report Card on Comprehensive Equity, are well known. Test score gaps in basic academic skills have been publicized widely, and several organizations have published compendia of health and child-rearing indicators. But this Report Card is the first attempt to have a group of national experts weight each available indicator by its relative importance. Using such weights permits the Report Card to make judgments about inequity in each broad goal area. The weights are also a guide to policy, pointing to areas where improvement might have the most impact in raising equity.

* An underlying assumption of the Report Card on Comprehensive Equity is that in our democratic society, government (state governments, in particular) have a responsibility to ensure that all children have a meaningful opportunity to enter young

adulthood with both adequate and equitable outcomes in the eight broad goal areas identified above. Governments should assign some of the responsibility for ensuring this meaningful opportunity to public schools, but should also assign appropriate responsibility to other institutions.

Thus, the Report Card is *not* about the outcomes of schools alone. Rather, it concerns the outcomes of all institutions of youth development, including schools. Both scholars and policymakers now broadly understand that even academic outcomes cannot be attributed to schools alone. The influence of schools in producing academic achievement is complemented by the influences of health care delivery systems, of the quality and availability of early childhood, after-school, and summer programs, of the quality of parenting, and of community environments (including safety, housing quality, racial isolation, and services such as public libraries). When this report uses terms like "institutions of youth development" we mean all of these educational, social, and economic institutions that affect youth outcomes.

Unfortunately, this underlying assumption is not usually reflected in educational or youth development policy. For example, awareness that approximately 25 percent of the black-white academic achievement gap is associated with differences in the health of black and white children, and in the health behaviors of their mothers, should lead policymakers to consider the relative value of placing health clinics in schools as a strategy for improving academic achievement. Similarly, awareness that approximately 14 percent of the black-white achievement gap is associated with differences in the residential mobility of black and white students should lead policymakers to consider the relative value of low-income housing support policies as a strategy for improving

academic achievement. Unfortunately, however, although aware that schools are only partially responsible for academic achievement and for racial achievement gaps, contemporary policymakers typically attempt to hold schools fully responsible.

But just as schools are not fully responsible for academic outcomes, they cannot be fully absolved of responsibility for other outcomes of education and youth development. Consider, for example, the goal of physical health. One aspect (among many) is the growing overweight and obesity status of young people. This troubling national development is characterized by racial disparities similar to those for academic achievement. It may result from deficiencies in the American diet, in the marketing of soft drinks and fast foods, in urban planning (less walking, less opportunity for safe outdoor play), and in recreational choices (computer vs. physical games). But it is also may be the product of school policies that prescribe inadequate time for active physical education and that fail to implement physical education curricula best designed to spur physical fitness. As with academic outcomes, schools and other institutions of youth development are jointly responsible for outcomes of good physical health. Similar co-responsibility of schools and other institutions of youth development apply to all the goals areas discussed in this report.

The Report Card on Comprehensive Equity highlights gaps in the eight broadly defined outcome areas of youth development. The policy implications of awareness of these gaps are not for school leaders alone. The policy implications are for state policymakers who, in seeking better youth outcomes, must balance their desires for school improvement with attention to all of the other institutions that also contribute to youth development.

There is at present little research, and certainly no conclusive research, that can guide policymakers in choosing the mix of school and other social and economic policies that will best provide a meaningful opportunity for all youths to enter adulthood poised for success. The Report Card on Comprehensive Equity implies no recommendations about the mix of such policies. Its sole purpose is to identify the black-white gaps in a broad range of outcomes. Such identification should spur policymakers to experiment with various mixes of school, social, and economic policies to narrow these gaps.

* The Report Card describes the black-white performance gap in the eight goal areas only at the national level. Typical discussions of the academic achievement gap attempt to measure it at state, district and school levels. But reasonably informative data on performance in the eight goal areas are presently available only at the national level, and even here, data are often inadequate. Without better data, policy makers cannot know which states perform relatively better than others. We recommend that national data collection be improved, and that such collection draw on samples that are large enough to support state-level conclusions about gaps in outcomes for each of the eight goal areas.

* It is conventional in policy discussions to consider inequity between blacks and whites, and inequity between Hispanics and non-Hispanic Americans, as being problems of an identical nature. The Report Card does not accept this assumption. This Report Card's concern is with performance in the eight goal areas of African-American youths, and how that performance compares with that of white youths. Seven generations after

the abolition of slavery, there is no shame of the nation greater than that of our failure to fully integrate black citizens into the mainstream of American society.

It is also important to measure the gap between the performance of children whose families immigrated to this country in recent generations, and that of children from fully assimilated families. Analysis of this gap can shed light on how well immigrants are being incorporated into our society. The ability of peasant and working class immigrants from Eastern and Southern Europe a century ago to benefit from schooling and then move into the mainstream has been an important success of American social history. We should try to understand whether educational and social institutions are providing immigrants today with similar opportunities.

But we can understand this only by measuring outcomes for children of families that have had a reasonable opportunity to assimilate. While appropriate education for recent immigrant children should be an important policy focus, it is different in kind from a focus on equity, a standard properly applied to children from families who have had a reasonable opportunity to assimilate.

Measuring equity for immigrant children presents enormous difficulties. The most commonly used classification for immigrants, "Hispanics" (or, sometimes, "Latinos") can be misleading. It combines immigrants, children of immigrants, and fully assimilated Americans with ancestry in Latin America or even Spain. It makes no distinction between peasant or unskilled worker immigrants, say from Mexico, and middle class refugees from Cuba or Central America. To evaluate whether children of recent immigrants are making progress comparable to that of children of immigrants a century ago, we need the ability to distinguish children whose parents were born outside the United States and

whose parents have less than a high school education. Otherwise, we may over- or underestimate assimilation and progress towards equity. Because data are presently unavailable to allow us to make such distinctions, at the present time we have made no effort to measure inequity between Hispanic-Americans and whites. In another aspect of our broader project (described below) we propose to correct these data limitations.

The Report Card defines the black-white gap by estimating the mean (average) black and white youth performance in national distributions of performance in each of the eight goal areas. The Report Card expresses these averages as the black percentile rank and the white percentile rank; the gap in each of these goal areas is the difference between these ranks.

Such comparisons of averages necessarily obscure great differences between individuals. There is a wide variation in performance among blacks and among whites in each of the eight areas. In each, although there is a black-white gap favoring white youths, there are some black youths who outperform average whites, and some white youths who underperform average blacks. In some cases, average black youths outperform average white youths on particular indicators, although this is not the case for any goal area as a whole.

Although each of the goal areas is, and has been an important objective of American educational and youth development institutions, each of the areas is not equally important. To help define the relative importance of each goal area, we conducted national surveys of representative samples of all adults, of school board members, of state legislators, and of school superintendents. We considered the results of these surveys,

together with our judgments and the judgment of experts, to weight each of the goal areas by their relative importance. The weights we recommend, and which the Report Card employs, are:

<u>Goal</u>	<u>Relative importance (percent)</u>
Basic Academic Skills in Core Subjects	21
Critical Thinking and Problem Solving	16
Social Skills and Work Ethic	14
Citizenship and Community	14
Responsibility	
Physical Health	9
Emotional Health	8
Appreciation of the Arts and Literature	7
Preparation for Skilled Work	11
Total	100

The Report Card on Comprehensive Equity concludes that the nation has a black-white performance gap of roughly the following magnitude:

* **Basic Academic Skills:** The black-white gap is about 29 percentile points. In a national distribution of achievement of basic skills by the time students are about 17 years old, black students are at the 31st percentile, and white students are at the 61st.*

* Here, and subsequently, percentile points are rounded to whole numbers, occasionally resulting in a gap that appears not to be identical to the difference in the whole-number percentile rankings of whites and blacks.

* Critical Thinking and Problem-Solving Skills: The black-white gap is about 31 percentile points. In a national distribution of achievement of critical thinking skills by the time students are about 17 years old, black students are at the 25th percentile, and white students are at the 56th.

* Social Skills and Work Ethic: The black-white gap is about 16 percentile points. In a national distribution of performance in social skills and work ethic by the time youths enter young adulthood, blacks are at the 41st percentile, and whites are at the 56th.

* Citizenship and Community Responsibility: The black-white gap is about 13 percentile points. In a national distribution of adolescents' and young adults' citizenship behavior and community responsibility, blacks are at the 42nd percentile, and whites are at the 55th.

* Physical Health: The black-white gap is about 7 percentile points. In a national distribution of readiness for lifelong physical health by the time youths enter young adulthood, blacks are at the 47th percentile, and whites are at the 54th.

* Emotional Health: The black-white gap is about 5 percentile points. In a national distribution of readiness for lifelong emotional health by the time youths enter young adulthood, blacks are at the 49th percentile, and whites are at the 54th.

* Appreciation of the Arts and Literature: The black-white gap is about 12 percentile points. In a national distribution of adolescents' and young adults' achievement in, appreciation of, and ability to participate in the arts and literature, blacks are at the 42nd percentile, and whites are at the 54th.

* Preparation for Skilled Work: The black-white gap is about 13 percentile points. In a national distribution of young peoples' preparation for successful careers if they are

not likely to graduate from college, blacks are at the 41st percentile, and whites are at the 54th.

* Overall Inputs: Influencing each of the eight performance gaps is an 18 percentile point gap in overall inputs (fiscal, school, family, and community) that contribute to a meaningful opportunity for performance in each of the goal areas. In a national distribution of access to such resources, blacks are at the 41st percentile, and whites are at the 58th.

As noted, the above summaries are based on available national data, which in every case are sparse. These estimates are the best that can be done with present data, although the nation could certainly collect better data on each of the broad outcome goals, were we to invest the necessary resources.

By applying weights reflecting the relative importance of each goal area to calculations of gaps in these areas, the Report Card on Comprehensive Equity concludes that schools and other institutions of youth development presently generate an overall black-white gap of about 18 percentile points. This performance gap results from the average black young adult being at the 38th percentile in this weighted distribution of performance in the eight goal areas, while the average white young adult is at the 56th percentile. The following table summarizes these results.

The Black-White Performance Gap in the United States Today			
	Blacks	Whites	Gap
Basic Academic Skills	31	61	29
Critical Thinking and Problem Solving	25	56	31
Social Skills and Work Ethic	41	56	16
Citizenship and Community Responsibility	42	55	13
Physical Health	47	54	7
Emotional Health	49	54	5
Appreciation of the Arts and Literature	42	54	12
Preparation for Skilled Work	41	54	13
Overall Black-White Performance Gap	38	56	18

These estimates are very approximate, given the inadequacies in data and the statistical manipulations necessary to make data from different sources comparable. It is not the intent of the Report Card to give a precise estimate of black-white inequity. Rather, it is the intent to emphasize, first, that measurement of comprehensive inequity must encompass all of the goals of education and youth development, not standardized test scores alone; and second, that imperfect though the data may be, measuring inequity across the many domains of education and youth development is both desirable and more feasible than is commonly thought.

The Report Card is part of a larger project, from which we anticipate additional, though complementary products. One will describe how, in the absence of precise quantitative measurement, schools might be held accountable for their contributions to adequate outcomes. Another will describe and offer cost estimates of a national data collection system of indicators, better suited than those available for this report, to provide meaningful assessments of the performance gap in all eight goal areas, for each

state where minority youths are present in significant numbers. This product is described below. A third will describe and offer cost estimates of the ingredients of several school and school-related programs that are likely to generate meaningful opportunity for adequate and equitable outcomes in each of the goal areas. We will present a preliminary version of these cost estimates at the Campaign for Educational Equity's annual symposium in the Fall of 2008.

Part II:
Summary of
*A Proposal for Expanded Data Gathering: How NAEP Can Monitor Equity on the
Full Range of Education and Youth Development Outcomes.*

Information already exists on whether states are raising the basic academic achievement of students in demographic subgroups. The federal government's National Assessment of Educational Progress, or NAEP (which calls itself "The Nation's Report Card"), shows how sub-groups in each state perform academically on a common scale, making comparisons possible.

NAEP is presently inadequate, however, for two reasons.

First, educational excellence results from a process that culminates when youths enter adulthood. Measurements of achievement at earlier ages can help guide instructional and social policy, but do not ultimately indicate whether states are making progress towards excellence with equity. Measurements of inputs can indicate whether states are making efforts to achieve equity, but not whether funds are being spent wisely to reach that goal. A valid report card on equity of outcomes should measure achievement no earlier than the end of twelfth grade (including results for youths who did not complete high school). To ensure that education and youth development institutions are delivering deep skills that are sustained, not skills that decay immediately after the end of formal schooling, outcomes should also be measured in young adulthood. Such reports of end-of-schooling and young adult achievement can be supplemented by data on intermediate outcomes and on inputs, but the only ultimately valid measure of excellence and equity is young adult achievement.

NAEP today assesses only in-school students in the fourth, eighth and twelfth grades. This is an abandonment of NAEP's original design of 40 years ago, when it assessed 17 year olds whether or not they remained in school, and young adults as well.

Second, NAEP provides only a distorted picture of achievement because it focuses primarily on basic academic skills and, to a lesser extent, on critical thinking. Indeed, NAEP's near-exclusive focus on academic skills also represents an abandonment of its original intent. NAEP's early design committee recommended in 1963 that ten subjects be included: the academic areas of reading, writing, mathematics, science, literature, and social studies, but also citizenship, art, music, and career and occupational development. John H. Fischer, then president of Teachers College, was a member of this exploratory committee. Lee Cronbach, one of the nation's most respected educational psychologists and a technical advisor to the committee, warned that if NAEP focused on basic skills alone, schools would take this as a cue to narrow their curricula. Thus, he urged, the national test should include the fine arts and should assess whether 12 year-olds understood the Bill of Rights. Other members of the committee urged that the national survey should also assess student attitudes and self-concept.

The early administrations of NAEP included a broader conception of achievement. For example, because of the emphasis which the original design team placed on behavioral outcomes, interviews about past or current actions were used in the citizenship exam at all age levels. Thus, one objective called for individuals to help others and to recognize ways for citizens to influence government action. To assess this at ages 13, 17 and 26-35, respondents were asked the following question in an interview format: “Suppose you and some friends were walking by a public park. As you went by, some

children of a minority group were stopped from entering the park by a man at the gate who told them, “this park is not for kids like you.” Would you feel that you should do something about it? What could you do about it if you wanted to?” Eighty-two percent of the 13 year olds, 90% of the 17 year olds and 79 % of the adults stated that they should do something in that situation and the majority of those who felt they should do something (over 80% for each group) could name at least one action they could take.

Adult respondents were also asked about their current civic participation levels during the interview. For example, they were asked if they had performed any unpaid civic activities such as doing volunteer work in a school or library or helping with a project to improve the community. Interviewers would follow up with respondents who reported “yes” to these types of questions by asking respondents to share more details about the types of activities he/she had performed and how often.

While questions that rely upon a respondent’s self-reported behavior can be subject to bias because respondents try to give the “socially correct answer” or they may inflate their actions in an effort to appear positively, NAEP interviewers were able to use follow up questions that asked for greater detail about an event, the number of times one had engaged in a specific activity or to recall and retell the most recent activity to verify answers. These follow up questions helped ensure that accurate responses were being given.

NAEP initially determined that “effective citizenship involves working well with other people. So to measure skills of interaction and willingness to participate and communicate, [NAEP found it] necessary to observe the behavior of respondents in group situations. Paper-and-pencil exercises are less expensive, and so are interviews, but

neither provide convincing measures of group interaction skills.” These group interactions were conducted at all age levels and trained observers recorded all interactions. Prior to the first official administration of the group interaction exercises, many cycles of development occurred. Observers received about three days of special training so that they could accurately and reliably assess the interactions.

Such assessments were possible even for the youngest age group assessed – age 9. At this age, groups of four students worked as a team to play “What’s in the Box?,” a game that challenged students to ask yes/no questions to determine what object was hidden in the box. Each game lasted for no more than 30 minutes and two teams raced to figure out the prize inside of the box first. Typical prizes were crayons or yo-yos. The game was designed to assess the ability of 9 year-olds to cooperate effectively in a group, display fairness, weigh alternatives, contribute good ideas to solve problems and the communicate effectively with others. Two observers systematically recorded all interactions of students and recorded when a student suggested a new question, gave a reason for or against a given question, encouraged the team to win, sought information or steered the task by organizing the group or suggested a change in procedure. The percentage of students who engaged in the above actions was reported in NAEP's citizenship report.

For older ages, similar skills were assessed, but a different group task was used. For the 13 and 17 year-old sample, a group of eight students were given a list of 12 public issues, asked to rank them in order of importance and to write a recommendation for addressing what they considered the two most important problems. Trained observers recorded the positive and negative behaviors of students as they engaged in this task.

NAEP then reported, for example, that only 4 percent of the 13 year olds defended the right of another group member to be heard or to hold a different opinion.

These early NAEP design elements - out-of-school samples, young adult samples, and performance assessments – were abandoned in the 1970s solely for reasons of cost.

In 1982, the National Assessment Governing Board commissioned former Secretary of Labor W. Willard Wirtz and a colleague, Archie Lapointe, to evaluate NAEP and make recommendations for its future. Wirtz and Lapointe concluded that "NAEP's coverage of the scope of education's impact should be broadened... [Confining NAEP] largely to the basic skills areas will virtually assure the Assessment's not playing a major role in informing the general public about the educational achievement picture." This recommendation was not followed.

Today, indicator data disaggregated by race, ethnicity, immigrant status, and family income, and by state, are almost entirely unavailable. Many surveys exist, but typically suffer from one of these limitations:

* Data have been collected for high school students, leaving out information on the achievement of youths who dropped out. This, for example, makes 12th grade NAEP an unsuitable source of information on equity in academic skills, especially because there is more likely to be missing data for disadvantaged youth who are more likely to have dropped out. The early designers of NAEP were sensitive to this problem, and insisted that NAEP assess 17 year-olds who were out of school; this practice was initially implemented, only to be abandoned in the 1970s for budgetary

reasons.* We propose instead a household-based survey, to fulfill the original purpose of NAEP, which was to assess all youth as they enter young adulthood.

* Data from several surveys are designed for making conclusions about national trends, but the samples are too small to support conclusions about the state-by-state performance of demographic subgroups. This is typical, for example, of many excellent surveys of young adults' physical and mental health conducted by the Centers for Disease Control. Twelfth grade NAEP is not administered at the state level, making it unavailable as a source for data on academic outcomes, even for students who remain in school for the 12th grade.

* Data on whether young people have achieved the outcomes sought by schools and other institutions of youth development are often not meaningful until several years after young people have left school. For example, if a goal of civics education is to prepare and motivate young people to vote, we cannot know whether this goal has been achieved until we determine whether they actually register and vote in the years immediately after their end of schooling. If schooling aims to develop an appreciation of literature and the arts, we cannot know whether this goal has been achieved until we determine whether young people read for pleasure or are engaged in the arts after schooling has ended. Likewise, the effect of physical education programs cannot be observed with accuracy until we know whether young adults exercise regularly.

There are, however, many national surveys which already include some, though not all of the features that should be included in a redesigned NAEP. These surveys were

* The in-school cost of the survey was \$5 while the out-of-school cost was \$50. (In current dollars, these amounts are equivalent to about \$25 and \$250, respectively.) The out-of-school NAEP survey was discontinued after the 1975-76 administration. Funding for NAEP declined from \$6 million in fiscal year 1973 to 3.9 million in 1980. In constant dollars, this was a 65 percent budget cut.

used by us in developing “*A Report Card on Comprehensive Equity: Racial Gaps in the Nation’s Youth Outcomes.*” Although it would not be cost-effective to attempt to simultaneously expand and co-ordinate these many surveys, much of the work in developing survey items and sample design has already been done. To produce a valid report card on educational excellence and equity, NAEP need not start from scratch.

A redesigned NAEP should collect original survey and assessment data on achievement in the eight goal areas. Sample sizes should be large enough to support reliable conclusions about the state-by-state achievement of white and of disadvantaged students. This requires data collected for whites, blacks, and Hispanic-American students. By "Hispanic-Americans," we mean "third- and subsequent-generation Hispanic immigrants": youth born in the United States of mothers who were also born in the United States.

NAEP should assess and survey 20 to 25 year-olds in each of these demographic groups. This age range is appropriate because the achievement of proficiency is effectively expressed at somewhat different ages for different goal areas. (For example, as noted, academic achievement could appropriately be assessed at age 17, but voting participation, involvement in the arts, and exercise habits, at later ages.) The 20 to 25 year-old range seems to be the youngest in which we can hope to capture data for each of the eight goal areas.

While it may today seem unusual to propose to assess educational equity by measuring the performance of young adults, such an approach was also part of the original NAEP design. NAEP originally surveyed all U.S. residents at ages 9, 13, 17, and "young adults," defined as 27 year olds (for practical reasons, NAEP approximated this

age group by including anyone between the ages of 25 and 35 who resided in a randomly selected household).

The rationale for surveying young adults in the original NAEP was partly to develop adequate vocational training programs and to provide information useful for improving adult education. But also important was an argument summarized in a 1964 memorandum by Peter H. Rossi, director of the National Opinion Research Center. It was prepared for John Gardner, president of the Carnegie Foundation which, at the request of U.S. Commissioner of Education Francis Keppel, had initially sponsored and financed the NAEP design.

Rossi argued that the long range goal of school systems is to produce an adult population with the "appropriate skills, knowledge and values" that were "sufficient to meet the needs of our society." Measuring such outcomes during the school years, or even upon graduation, was of limited value because there may be "school systems which produce graduates whose induced characteristics decay at a faster rate than those of other school systems." Perhaps, Rossi reasoned, "there are differences in the decay curves of skills, knowledge and values which can be related to the characteristics of school systems," and if so, a national assessment designed to measure educational progress would fail to distinguish successful from unsuccessful education and youth development systems if the assessment only gathered data from students still in school or even upon graduation.

NAEP initially adopted Rossi's recommendation, and its initial survey included a young adult sample. This young adult assessment was also dropped in the 1970s for

budgetary reasons, but the importance of ensuring that outcomes were measured for their stability through young adulthood was not repudiated.*

We think it reasonable for the federal government (through an expanded NAEP) to measure equity for the nation and for the 36 states where there are concentrations of disadvantaged youth – 28 states (including the District of Columbia) where black-white equity can be measured, and 16 states where equity for Hispanic-Americans can be measured. There are both policy and practical reasons for limiting the Report Card to these states. From a policy perspective, states with large numbers of disadvantaged students face problems in achieving equity for them that are qualitatively different from states with few such students. Concentrated disadvantage is itself one of the most powerful risk factors for children. Policies to overcome such concentration necessarily will differ from policies designed to achieve equity for disadvantaged students who are small minorities in their communities and states. From a practical perspective, accurate measurement of equity is possible only for states with sufficiently large populations of disadvantaged students from which to draw samples. The requirements to locate samples of very small populations are cost-prohibitive. Further, even if cost were not a concern, reliable estimates cannot be drawn from data collected where samples are too small.

The District of Columbia is a unique case. Although its black youth population is large enough to sample, its white youth population is not. Therefore, although we recommend that NAEP collect data on black youth in the District of Columbia, for purposes of measuring equity, data on the outcomes of these youth can only meaningfully

* The Wirtz-Lapointe report, referenced above, also recommended that NAEP return to assessing out-of-school 17 year olds and young adults. This recommendation, too, was not followed by NAGB.

be compared to the outcomes of the national white young adult population, not with whites in the District alone.*

Some young adults who will be surveyed in such a design will be sampled in states different from those in which they spent their youth. Data collected from these young adults should be assigned to the state where they spent two of the following three points: birth, third grade, and completion of schooling. Young adults who were not in the same state for two of these three points should not be included in the data collection. In the case of college students living away from home, NAEP should include them in the survey if the household member initially contacted can provide a student's location. Preferably, the sample development should take place during a school vacation period when college students are more likely to be at their permanent residences.

We estimate that a full cycle of surveys and assessments to generate such data would cost a minimum of \$35 million. An additional survey of businesses that employ 20 to 25 year-olds would add another \$4 million in cost. If, to measure state-by-state progress, such surveys and assessments were conducted on an ongoing basis every three years, the average annual cost would be approximately \$13 million.

There will be additional administrative and development costs, including those associated with the organization and convening of expert panels to assign within-goal weights. Thus, we estimate the total cost of such a survey to be initially \$45 million, with ongoing costs of \$13 million annually beginning with the fourth year following the initial survey.

* As noted below, residential moves of young adults out of the District and into the Washington-area suburbs may present an obstacle to including a representative sample of District youths in the survey.

A NAEP so re-designed could play an important, but limited role in national and state educational and youth development policy. By identifying states with greater overall inequity in the eight goal areas, it can spur those states to adopt policies to improve their performance. States should be able use the information on young adult achievement provided by such data to judge whether resources should be increased, or whether resources should be utilized differently.

Even redesigned, however, NAEP cannot provide guidance regarding which policies states should adopt to pursue the objectives of greater excellence with equity. If a state, for example, is producing disadvantaged youth in poor physical health, with greater risk for premature death, NAEP itself cannot contain information to guide state policy makers about whether they should require more physical education in the school curriculum, provide better after-school and summer options, or regulate the nutritional content of foods available to disadvantaged youth. If a state is producing disadvantaged youth with inadequate critical thinking skills, NAEP itself cannot contain information to guide state policy makers about whether they should revamp the school curriculum or improve the availability of high quality early childhood programs. Likewise, if a state is producing youth who excel, for example, in civic responsibility, this information may help policy makers in other states to identify programs that are succeeding, but cannot itself confirm their effectiveness.

Nor can a re-designed NAEP enable comparisons of achievement between schools or school districts within a state. The cost of enlarging samples to generate district- or school-level results would be prohibitive. It might not, however, be prohibitive to expand a redesigned NAEP to include results for a few of the nation's

largest urban school districts, as NAEP presently does for children's academic skills in its Urban Assessment.

With NAEP so redesigned, states may be inspired to hold schools and school districts accountable for the broader range of outcomes that Americans want from their schools and other institutions of youth development. Then, the expense of necessary surveys and assessments for district- and school-level results will be borne by state governments. This is as it should be. But at a time when no data are available anywhere that would permit us to understand the relative success of states in achieving educational excellence with equity, a redesigned NAEP can be an important first step in the development of adequate theories of accountability nationwide.