National Institute
for Science Education
University of Wisconsin-Madison • National Center for Improving Science Education

Research Monograph No. 9

Reaching Equity in Systemic Reform: How Do We Assess Progress and Problems?

Jane Butler Kahle

Funded by the
National Science Foundation

National Institute for Science Education (NISE) Publications

The NISE issues papers to facilitate the exchange of ideas among the research and development community in science, mathematics, engineering, and technology (SMET) education and leading reformers of SMET education as found in schools, universities, and professional organizations across the country. The NISE Occasional Papers provide comment and analysis on current issues in SMET education including SMET innovations and practices. The papers in the NISE Research Monograph series report findings of original research. The NISE Conference and Workshop Reports result from conferences, forums, and workshops sponsored by the NISE. In addition to these three publication series, the NISE publishes Briefs on a variety of SMET issues.

The research reported in this paper was supported by a cooperative agreement between the National Science Foundation and the University of Wisconsin–Madison (Cooperative Agreement No. RED-9452971). At UW–Madison, the National Institute for Science Education is housed in the Wisconsin Center for Education Research and is a collaborative effort of the College of Agricultural and Life Sciences, the School of Education, the College of Engineering, and the College of Letters and Science. The collaborative effort is also joined by the National Center for Improving Science Education, Washington, DC. Any opinions, findings, or conclusions are those of the author and do not necessarily reflect the view of the supporting agencies.

Research Monograph No. 9

Reaching Equity in Systemic Reform: How Do We Assess Progress and Problems?

Jane Butler Kahle

National Institute for Science Education
University of Wisconsin–Madison

April 1998

About the Author

Jane Butler Kahle, Condit Professor of Science Education at Miami University, Oxford, Ohio, has directed 32 externally funded research projects. Her current research involves testing a model for gender research in science education as well as assessing the effects of systemic reform from an equity perspective. During the past seven years, she has been principal investigator of Ohio's State Systemic Initiative; currently she has a research grant to continue the assessment of that reform. She has published widely and has received several national awards and international fellowships for her scholarship.

Kahle has been president of the National Biology Teachers Association and the National Association for Research in Science Teaching. In addition, she has chaired the Board of Directors for the Biological Science Curriculum Study (BSCS), the Gender and Science and Technology Association, and the Committee on Equal Opportunities in Science and Engineering at the National Science Foundation. Currently, she serves as chairperson of the National Research Council's Committee on Science Education, K–12, and as a member of the Advisory Committee for the Directorate of Education and Human Resources at the National Science Foundation. During 1996–97, Kahle was a Fellow at the National Institute for Science Education, University of Wisconsin–Madison, where she developed a model for assessing equity in systemic reform.

Acknowledgements

The author wishes to thank Norman Webb and Andrew Porter for their thoughtful reviews of this paper.

An earlier version of this paper was published as Equitable Systemic Reform in Science and Mathematics: Assessing Progress in the *Journal of Women and Minorities in Science and Engineering* (1998, *4*, 91-112).

Contents

Abstract

This paper provides an analysis of educational equity in science and mathematics and proposes a practical way to assess equity in systemic reform. A synthesis of major national and international studies as well as a comprehensive review of the literature are used to suggest a way to assess when educational systems, particularly those involved in systemic reform, are moving toward equity. First a definition of equity and of equitable education is developed. Next, based on an analysis of NELS:88, High School and Beyond, and TIMSS, indicators of equity are proposed. The efficacy of those indicators is analyzed by a review of the pertinent literature. Last, an *equity metric* is developed.

Using a research paradigm, the paper builds a practical model that is applicable for districts, schools, and classrooms across this country. Key indicators, applicable across many types of educational systems, are enrollment and achievement in eighth-grade mathematics, quality of the content and instruction of science and mathematics courses, narrowing of achievement gaps among subgroups of students, and changes in teaching practices to align them with the mathematics and science standards. Variation in the *equity metric* is described for several types of school districts.

Introduction

> Any serious attempt at change in U.S. science and mathematics education must be deeply structural. The fundamental problem is not a conglomeration of individual problems. Any effective reform in this context will necessarily be systemic—affecting several parts at once.
>
> Schmidt, McKnight, & Raizen, *A Splintered Vision*

To address effective reform, individual schools as well as districts, states, regions, and the nation have become involved in systemic reform. The driving force behind the current reform movement, as with all other reforms of education in this century, has been the need to substantively improve public education in our country so that we remain competitive—economically, scientifically, and technologically—with other developed nations. Each time another country achieves a victory in one of those arenas, we look to education to even the score. Increasingly, as our population has become more diverse and as the underrepresentation of whole groups of students in science and mathematics has become more visible, we have come to understand that educational reform must be both systemic and equitable. That is, the reform must address multiple parts, while increasing the access, retention, and achievement of all students in high quality science and mathematics programs. Curricula must change so that varied interests are represented. More effective ways of organizing classrooms and schools and of providing instruction must be implemented, and assessments need to include multiple ways of demonstrating learning and competencies. In addition, policies that determine both the quantity of courses and the quality of the educational experience (teacher qualifications, instructional resources, and academic tracking, among others) must be reviewed and changed to ensure equitable reform.

However, complex factors impinge daily on any given child's readiness to learn, factors that are beyond the scope of any educational system. How do we ensure—this time—that the reforms are both systemic and equitable? What kind of indicators suggest that a classroom, school, or district is becoming more equitable? And just exactly what does "equitable" mean?

A model for assessing progress in achieving equitable systemic reform in science and mathematics education is suggested in this paper. It does not provide a simple solution, rather it proposes an evolving and complex mechanism. Further, although examples are provided, it does not supply a single solution. Each state, region, district, school, or class that proposes to meet the needs of all students equitably will need to select and use the indicators that are most appropriate for its situation and conditions. Finally, the model does not provide a short-term solution. As progress is made, other groups of children may become underrepresented so that different indicators and criteria will be needed.

Although the model for assessing progress toward achieving equitable systemic reform is not a simple, single, or short-term solution, it is grounded in research. Data have been triangulated from large national studies, specifically the National Education Longitudinal Study of 1988 (NELS:88), High School and Beyond, and the Third International Mathematics and Science Study (TIMSS)[1] and have been synthesized with findings from the research literature. Measuring progress toward equity in science and mathematics education suggests the need for an *equity* *metric*—that is, a metric that allows the monitoring of progress over time toward, or away from, equity. In the rest of this paper, the need for an equity metric, the principles on which it is founded, the indicators that define it, and the measurement and psychometric issues that affect it are explicated. Last, a sample equity metric is provided.

Need for an Equity Metric

> The notion that equity in education depends on the strength of the linkage among beliefs, opportunity, and achievement has been a key assumption of systemic reform. Systemic evaluation should build a growing understanding of this model and should identify key pressure points in the system with regard to equity.
>
> Heck & Webb, *Purposes and Issues*

A brief review of student learning outcomes in mathematics and science clearly indicates that student achievement may be categorized by membership in specific subgroups. Although National Assessment of Educational Progress (NAEP) data are not disaggregated and reported by gender **and** by race, they provide a broad look at achievement by either race/ethnicity **or** by gender. NAEP results, which are reported every four years, approach a national report card for American children and allow tracking of progress over time. Briefly, from 1977 to 1992, there was a slight decline in the gap that favors boys on anchor items (ones that were stable across the years) in science for 17-year-olds. That is, during that period, girls' achievement levels rose more than did those of boys. The change in achievement for students in different ethnic/racial groups was mixed, depending on test subject matter and student age. However, results indicated that the gap between the percentage of white students and African American and Hispanic students scoring at selected anchor points decreased for mathematics and, to a lesser extent, for science between 1977 and 1990. Between 1990 and 1992, the science gap between whites and the other two groups increased on anchor items.[2]

NAEP data also show that in 1992 about the same proportion of white, African American, and Hispanic high school graduates earned credits in biology and introductory algebra, but that significantly higher proportions of white graduates completed courses in chemistry, physics, geometry, advanced algebra, and trigonometry. When course enrollment patterns are reviewed by gender, the only differences are found in physics (taken by 28% of males and 21% of females) and advanced placement physics. Contrary to popular belief, girls' and boys' enrollments in trigonometry (21%) and calculus (8%) are equal (National Science Foundation [NSF], 1996).

The consistent differences in enrollment patterns and achievement by subgroups of students were a driving force behind the NSF's 1991 initiative to address the systemic reform of mathematics and science education. The Foundation defines systemic reform in the following way:

> "Systemic reform" is a process of educational reform based on the premise that achieving excellence and equity requires alignment of critical activities and components. It involves a change in infrastructure as well as outcomes. Central elements include:
>
> - high standards for learning expected from all students;

- alignment among all the parts of the system—policies, practices, and accountability mechanisms;
- a change in governance that includes greater school site flexibility;
- greater involvement of the public and the community;
- a closer link between formal and informal learning experiences;
- enhanced attention to professional development; and
- increased articulation between the precollege and postsecondary educational institutions. (NSF, 1996, p. 5)

Although NSF provided a broad definition of systemic reform, it did not define what the "system" was. In fact, it defined at least three levels of "systems." Initially, the system was identified as a state's educational system; next, single school districts (albeit large ones) were the system for reform; and, third, the system was defined as representing multiple schools, districts, and states serving specific populations of students with certain identifiable characteristics living in specific areas (Appalachian area or Native American reservation schools and colleges, for example). In this paper, the district will be considered the system in systemic reform; that is, the district is the level at which the question, "How do we know when a system is moving toward equity?" is addressed.

Four years into its systemic reform initiative, NSF identified components that drive systemic reform and requested grantees to assess progress against them. The components, called drivers, include:

- implementation of comprehensive, standards-based curricula, instruction, and assessment in every classroom, laboratory, or other learning experience;
- development of a coherent, consistent set of policies that supports high quality science and mathematics education for all students, continuing education for all teachers of science and mathematics, and administrative support;
- convergence of all resources that could be used to support science and mathematics education into a unitary program to upgrade science and mathematics education for all students;
- collaboration and broad-based support from parents, policymakers, institutions of higher education, business and industry, foundations, and other segments of the community;
- accumulation of a broad and deep array of evidence that the program is enhancing student achievement; and
- improvement in the achievement of all students, including those historically underserved.

Many of the NSF drivers address the issue of equity in systemic reform. In fact, they address issues of social justice. However, they do not suggest how to measure progress toward that goal, nor do they address the complex issue of cultural values.

Others have addressed the issue of systemic reform and how it relates to and affects assessment and evaluation. Romberg (1995) suggests that evaluation of systemic reform must assess a range of variables, including mathematics and science content, teaching, technology, policy, grouping of students for instruction, pedagogical decisions by teachers, student pursuits, and student performance. However, that range of variables does not begin to assess all of the central elements of systemic reform identified by NSF. As discussed below, the disconnection between what have been defined as elements in systemic reform and what is feasible and possible to assess is one of the challenges in evaluating systemic reform and in developing an equity metric.

Equity—An Elusive Target

> Words such as equity are sliding signifiers. They do not have an essential meaning . . . but are defined by their use in real social situations with real relations of power.
>
> Apple, *Taking Power Seriously*

A first step in developing an equity metric is to agree on a working definition of equity that is broad enough to encompass reasonable subgroups (e.g., minority cultures, groups with exceptional needs), flexible enough to accommodate the dynamics of changing demographics, meaningful enough to ensure viability in the political-social environment, and sensitive enough to accommodate a viable assessment of equity (Strategies for Evaluating Systemic Reform, 1996). A broad definition of educational equity involves equitable policies and practices, equitable distribution of resources, equitable cultural and peer attitudes, equitable cultural identities, and equitable identification and/or categorization of groups. However, developing a definition is compounded by a blurring, or changing, of what are identifiable groups. As Clune (1996) explains,

> An argument could be made that the social consensus underlying the meaning and purpose of equity is changing in ways which are appearing as "complications" in the field (of practitioners and policy makers). The old reliable categories of Hispanic and Black are beginning to seem conceptually inadequate for new realities. You have [scholars] suggesting that Hispanics be thought of as "Latino" and broken down into categories which are more indicative of "real need." You have images of "new immigrants" creeping into the discourse, without any clear definition of the details. The category of African American includes middle-class well-educated Blacks as well as desperately poor, socially disorganized people in ghettoes. You have the impossibly complicated racial identities of third and fourth generation people who think of themselves as "mixed race." (p. 1)

The assessment of systemic reform not only requires a broad and flexible definition of equity, but it also involves a reconceptualization of what can, and should, be measured. Clune (1993) has argued that equity in education should move toward the equalization of outputs, rather than the equalization of inputs. Porter (1994) extends the discussion of inputs and outputs and concludes that both student and school outputs must be measured.

> The history in the United States of holding students accountable for their achievement but not holding schools accountable for what they produce is curious. . . . Student achievement is a function of not only what the student puts forth, but what the school puts forth as well. To make matters even more complicated, student achievement is a function of what the home and community provide by way of out-of-school support. (p. 493)

He favors holding schools accountable for what they produce in terms of disaggregated student achievement, stating that the "goal is to get students from any identifiable subgroup to have an adequate distribution of achievement, including individuals who reach the highest levels" (p. 499). By holding schools accountable, Porter maintains that one can assess the value added; that is, "the increase in student achievement during the period of time that the student attends a particular school" (p. 494). In systemic reform, systems (districts or states) as well as schools and students are held accountable, and the value added applies to the student's tenure in a particular system. The assessment of equitable systems that follows is focused on outputs, uses the national mathematics and science standards to describe achievement goals, and has indicators that address equity at the student, teacher, school, and system levels.

As shown in Figure 1, equity at the system level has at least three dimensions. First, it involves a consideration of the *resources* available: What are a given family's resources? How does a specific group identify or characterize itself in relation to education? and What resources are available in the community? The system's *educational plan and practices* are a second dimension: What is the quality of the curriculum and of teacher preparation? Does the treatment of students differ across subgroups; if so, what is the evidence? Do students, parents, teachers, and administrators hold similar goals for subgroups of students? The third dimension of equity at the system level is student *outcomes*: Do the achievement levels or course enrollment patterns of students differ by subgroups? These three dimensions and the identified components provide a scaffold for building an equity metric based on research-validated indicators.

Building upon the three dimensions in Figure 1, we can define an equitable system as one in which identifiable subgroups of people do not experience systemic discrimination in process, in opportunities, or in negative outcomes without an ethically sufficient reason.[3] An equitable system serves the mathematics and science learning needs of each student and communicates this expectation to all members in the system. An equitable system has the means to reflect on the achievement of goals, and it has a mechanism for holding staff and students accountable for the achievement of equity goals. It promulgates flexible policies and practices that take into account the needs of each student regardless of race/ethnicity, gender, disabilities, family background and resources, language, and community, and/or school wealth. Equitable systems should provide agreed-upon opportunities and support for all students, teachers, families, and community members. Because equitable systems will differ from each other, any metric to assess equity must have multiple components and a range of measures.

In summary, an equitable system is:

- one in which all children have the opportunity to achieve to their fullest potential or to the levels specified in the system's performance standards;[4]
- one that is committed through its allocation of resources to the equitable achievement of all culture- and gender-based student populations;

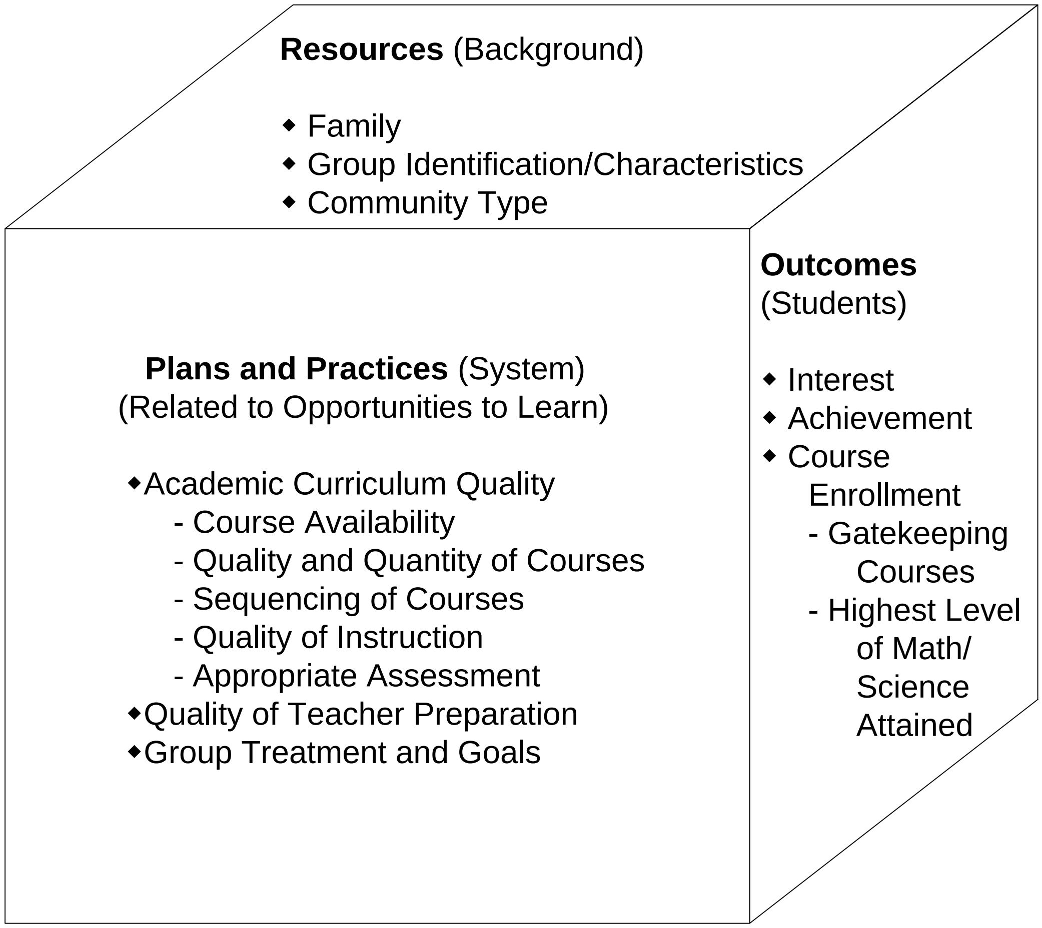

Figure 1. *Dimensions of equity*

Source: Anderson, B. (1996, November 20-21) in a paper presented at the meeting on Measuring Equity in Large Systems, University of Wisconsin–Madison, National Institute for Science Education, Strategies for Evaluating Systemic Reform Project.

- one in which participation of diverse groups, particularly those groups traditionally under-represented in the system, is expected and facilitated;
- one that is accessible; for example, sensitivity to individual variation is considered; and
- one that has policies and procedures established and followed for distributing and utilizing resources in ways that narrow any identified differences between subgroups.

The above definition suggests general areas for assessment. However, because of the complexity and comprehensive nature of an equitable system, assessing one requires ongoing monitoring. Instead of thinking of a *one-time* measure, repeated measurements of movement toward equity are needed. Some variables, such as enrollment and achievement patterns, will require continuous assessment; others, such as scheduling policies, may need to be evaluated less frequently. The next section considers how opportunities to learn and indicator studies contribute to the articulation of an equity metric.

Opportunities and Indicators—Defining an Equity Metric

> Observers who attribute minority students' low achievement solely to society's ills tend to dismiss the very important influences of schools and teachers on learning.
>
> Stevens, *Closing the Achievement Gap*

As discussed, NSF's definition of systemic reform as well as the drivers with which it intends to assess those reforms go well beyond the individual classroom, teacher, or school. Yet, the impact of the reforms on student learning is greatly influenced by individual teachers in separate, and somewhat autonomous, buildings. This dichotomy presents an additional challenge for the articulation of an equity metric. A review of the evolving concept of opportunities to learn and a consideration of how opportunities to learn help define indicators of equity follow.

The concept of opportunities to learn has changed dramatically in the last few years, particularly with the release of the National Council of Teachers of Mathematics' (NCTM, 1989) *Curriculum and Evaluation Standards for School Mathematics* and the National Research Council's (NRC, 1996) *National Science Education Standards*. Opportunities to learn originally focused on specific conditions of teaching and learning that generated high academic achievement; that is, how teachers taught students and what type of access to information and resources students needed to learn the curriculum for their age and grade level. In the 1970s the issue of social ills led to a pendulum swing away from defining learning opportunities in terms of school- or classroom-based opportunities to a concern about a variety of community- and home-based conditions that affect children's learning opportunities. Stevens' warning in the above quote suggests that a middle ground is needed. That is, any assessment that attempts to measure school- or classroom-based opportunities to learn also must consider the variety of home and community factors that impinge on a system's movement toward equitable education for all students.

Stevens (1996) expands the definition of opportunities to learn to include instructional quality, school and family support, student classroom performance, and student classroom behavior. She bases her expanded definition on responses from research directors, representing 91 public school districts, who were asked to identify four variables to explain differences in student achievement (Stevens, 1993). In rank order, the following variables were identified: (1) content coverage (how much of the curriculum is studied), (2) content exposure (the depth to which the curriculum is studied), (3) content emphasis (particular components emphasized in the curriculum), and (4) quality of instructional delivery (the way the curriculum is taught). She also identifies teacher professional characteristics (for example, development of content knowledge as well as knowledge and use of diverse types of assessment) and school support systems that affect student learning opportunities (e.g., amount of time allotted for teacher collaboration and time for teachers to become part of an *active* professional community).

Porter (1993) has suggested that opportunity-to-learn standards can be used to provide a basis for school-by-school accountability as well as an indicator system that describes the extent to which the implemented curriculum is consistent with content standards. However, he cautions against using opportunity-to-learn standards to establish school accountability, because that use could shift a reform's emphasis from student outcomes to school inputs. On the other hand, because student outcomes are related to the quality of the implemented curriculum, the actual learning opportunities that occur within classrooms need to be assessed. His research has shown that teacher logs provide an effective way to assess the nature of the implemented curriculum. As discussed later, ways to measure the quality as well as the quantity of science and mathematics courses are needed in an equity metric.

For this paper, three large databases—NELS:88, High School and Beyond, and TIMSS—have been synthesized to assess factors beyond the characteristics of teachers. All three databases are national in scope, and two of them, NELS:88 and High School and Beyond, report longitudinal data that facilitate identification of change over time. Using recent analyses of these databases as well as NSF's indicators of quality mathematics and science education (NSF, 1996), indicators were identified to assess progress toward equity. Answers to three questions were sought: first, what have been indicators of **inequality** between or among subgroups of students (e.g., unequal participation or achievement); second, when or where are the indicators evident in the educational system (e.g., grade level or type of course); and, third, are they found in more than one database? If evidence of inequality was found for an indicator in more than one database, that indicator was included; for example, both NELS:88 and High School and Beyond found that enrollment in eighth-grade algebra was unequal across subgroups of students, with whites and Asians enrolled more frequently than African Americans and Hispanics.

Next, the identified indicators were sorted by grade levels, including the grade level of students when the information was collected and the grade level when enrollment and participation were critical for continued access to and/or progress in science and mathematics. This sorting suggested leverage points in the educational system. Leverage points were selected for the developing metric if a consistent pattern was found for a grade or age level in more than one database; for example, eighth grade was selected as a leverage point because it is the time for enrollment in algebra as well as for selection of an academic program. The leverage points related to critical times in a child's education; that is, times when educational systems routinely gather data concerning specific placement (e.g., general mathematics or algebra) and performance (e.g., standardized achievement test results, high school graduation). Because the specifics are critical to a student's progress, they become points for leveraging equity in a system. The leverage points identified were preschool and fourth, eighth, tenth, and twelfth grades. Last, the identified indicators were aligned with the appropriate leverage points.

As the synthesis proceeded, indicators of general reform were needed. Using the same criteria and databases, overall indicators of systemic reform in science and mathematics were added to the developing metric. Because most of the indicators identified in the synthesis addressed middle and high schools, the literature was searched to identify appropriate indicators for elementary education and to verify the ones identified from the large databases. All of the indicators selected (shown in Figure 2 and discussed below) are based on empirical evidence that they affect the access, retention, and/or achievement of students by identifiable subgroups.

Figure 2 expands on the dimensions in Figure 1. Indicators of equitable *Plans and Practices* are found in *Access* and *Retention*, while indicators of *Outcomes* are delineated under *Achievement*.

Indicators	Leverage Points				
	Pre-K	Grade 4	Grade 8	Grade 10	Grade 12
ACCESS					
Home Resources	●	●	●	●	●
Minutes/Day of Math/Science		●			
Enrollment in Algebra/Geometry			●		
Enrollment in Calculus/Physics					●
Adademic Program				●	●
Expected Academic Program			●		
Limited English Proficiency		●	●	●	●
Quantity/Quality of Math/Science Courses			●	●	●
RETENTION					
Instructional Quality	●	●	●	●	●
Teacher Expectation/Behavior	●	●	●	●	●
Teacher Morale		●	●	●	●
Teacher/Student Attitudes and Beliefs			●	●	●
Learning Behavior		●	●	●	●
Critical Mass			●	●	●
Student Mobility		●	●	●	●
Out-of-School Experiences	●	●	●		
ACHIEVEMENT					
Increase in Eighth Grade Math Achievement			●		
Increase in Graduation Rates					●
College/Labor Market Performance					●
Decrease in “Gap”		●	●	●	●
Meet Local College Admission Requirements					●
OVERALL					
Equity Plan		●	●	●	●
Plan Implemented		●	●	●	●
Teacher Mobility		●	●	●	●
Increase in Availability of Advanced Math/Science Courses			●	●	●
Increase in Math/Science Graduation Requirements					●
Incentives for Change/Equity		●	●	●	●
Quality of Professional Development		●	●	●	●

Figure 2. *Research-validated indicators of equity*

Sources: Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996; Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1996; Fennema, 1990; Kahle, 1996; Lane, Wang, & Magone, 1996; NCES, 1995; NSF, 1996; Oakes, Ormseth, & Campbell, 1990; Olson, 1996; Porter, 1993; Schmidt, 1993; Secada, 1995; Stevens, 1996; Wolk, 1996.

Further, the educational level at which each indicator is commonly assessed (and, consequently, the level at which it becomes a leverage point for moving toward equity) is provided. In addition, indicators of a system's progress toward equitable education are included under *Overall*.

A primary caution about the approach used is that opportunity-to-learn analyses and indicator studies do not explain many of the gender differences found. As stated earlier, girls and boys enroll in equal numbers in algebra, biology, calculus, chemistry, and trigonometry; further, on average, girls achieve higher grades than boys in those courses. However, the enrollment patterns in physics are not equal, suggesting that neither course enrollment patterns nor achievement in science and mathematics predicts girls' enrollment in physics (National Center for Education Statistics [NCES], 1997).

Rather, more subtle influences for which we do not yet have adequate, or standard, measures seem to affect girls' participation in science and mathematics. More sensitive indicators as well as varied methodologies for gathering data may be required to assess gender equity. Continued surveillance will be needed to ensure that equity indicators are sensitive enough to assess factors affecting girls as well as emerging subgroups of students.

In developing an equity metric, the intent has been to identify indicators that:

- are sensitive to diversity among groups;
- are used to inform action, not only to define the present state;
- are flexible, because not all metrics are relevant to all parts of the system;
- distinguish between opportunity, accessibility, and participation;
- are directed toward leverage points in the system; and
- are feasible to use (i.e., affordable).

Once appropriate indicators have been identified and commonly accepted, educational systems can use them to design an equity metric. Indicators with the characteristics listed above move beyond a statistical analysis of student achievement as the primary measure of equity in a system. At the broadest level, they may be used to answer the question, Does the system address the indicators for equity?

The synthesis across the large databases, NSF's indicators, the research literature, and the assessments of several systemic initiatives suggested indicators of movement toward equity. In addition, it revealed several past indicators that may no longer be applicable, notably, level and type of teacher education. For example, the recent synthesis by Wolk (1997) as well as earlier studies by Oakes, Ormseth, and Campbell (1990) identified teacher qualifications (undergraduate degree in science or mathematics and years of experience) as well as type of certification as indicators of student learning outcomes. Further, because teacher qualifications differed between schools serving primarily minority versus primarily majority students, they could be used as an indicator of equity. However, recent analysis of NELS:88 data (NCES, 1995) indicates that teachers of minority students are not necessarily less well prepared than teachers of white students in terms of type of certification, number of years in teaching, or their own educational level. There are no significant differences in science, and the only difference in mathematics is in percentage of certified teachers of Native American students, compared to all other groups. Although the current analysis suggests dropping teacher qualifications as a factor in the developing metric, future data may warrant its inclusion again. Or, the nature of indicators that address teacher qualification may need to be refined. Instead of using certification, experience, and a bachelor's degree, indicators of the quality of the teacher preparation and programs may be needed. For example, more useful indicators may be number of credits in science and mathematics, evidence of advanced as well as introductory science and mathematics courses in the undergraduate program, length and quality of practicum or intern experience, and certification by the National Science Teachers Association or the National Board for Professional Teaching Standards. Today, certification is based on the number of academic credits in physics, for example; tomorrow the quality (laboratory-based, extended inquiries, etc.) of the physics course may be the indicator. The appropriateness of indicators will evolve as systems move toward equity.

Further, some factors may not apply to all subgroups; for example, only Hispanic and Native American children may be identified as subgroups when the indicator *Early Start of Schooling* is applied (NCES, 1995). Recent reports have provided conflicting findings concerning the efficacy of the rate of high school graduation as an indicator that distinguishes among subgroups of students.[i] Therefore, the indicator suggested for the equity metric is *Increase in Graduation Rates*. The intent is to assess graduation rates by subgroups of students and to follow those rates for several years.

Some of the indicators, identified in one or more studies, need further elucidation. For example, *Limited English Proficiency* is included in the model because access to science and mathematics education is restricted when language is a barrier, if only because the time spent in language instruction is time not available for science and mathematics instruction. The TIMSS database provides an interesting look at *Home Resources*, one of the most difficult indicators to measure. TIMSS measured the indicator *Home Resources* by the presence of a dictionary, access to a home computer, educational level of either parent, and table or desk for student's own use. Two of those four, the presence of a table or desk for the student's own use and a computer in the home, were linked to student achievement in many of the 41 countries (including the U.S.) in the TIMSS study (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996). Other evidence of *Home Resources* that separated groups in the large databases was attendance at a nursery or preschool (NCES, 1995). It is relatively easy to measure language, presence of a desk or table, and attendance at preschool, and they will provide added dimensions to the metric as well as meet the criterion of specificity to subgroups. They are included as part of the indicator *Home Resources* in Figure 2.

Although a variety of student characteristics affect retention in mathematics and science, two major categories are found in Figure 2. First, *Student Attitudes and Beliefs* addresses the documented decline in positive attitudes between fourth and twelfth grades. It is easy to measure and it is an indicator that also focuses on gender equity, because the decline in attitudes is greater for girls than for boys. Second, the indicator *Learning Behavior* is a multidimensional one that includes absenteeism and tardiness (which are easy to measure and indicate the degree of student engagement in learning) as well as the priority students place on learning and the amount of competition students face for grades (increasing competition correlates with decreasing achievement for non-Asian minority groups). Both were found in more than one database and the research literature, and both are included in the model equity metric.

One of the most interesting indicators is *Quality/Quantity of Math/Science Courses*. Evidence for this indicator comes from Schmidt (1993), who suggests that the quality of the content taught differs across courses. He identified six patterns (algebra, arithmetic, geometry, enriched arithmetic, potpourri, and remedial) in the mathematics curriculum in terms of the number of hours allocated to different topics (algebra, geometry, ratios, fractions, etc.). Then, he ascertained whether the amount of time spent on a topic differed from pattern to pattern. For example, did the amount of instructional time allocated to algebra topics differ when those topics were taught in the context of an algebra curriculum (or pattern) or when they were taught within another pattern such as the enriched arithmetic curriculum? He found that the amount of time did not differ markedly (26.3 hours in algebra versus 21.8 hours in enriched arithmetic) and that the topics covered (e.g., formulas and equations) were essentially the same. However, different achievement results were found for students enrolled in algebra and those in enriched arithmetic. Schmidt hypothesized that those differences in achievement were due to the quality of content (depth of study) and instruction (extended problem solving) in the two curriculum patterns (algebra and enriched arithmetic).

Likewise, Porter's (1993) analysis of teacher logs suggests that quality of content includes both amount of coverage (time actually spent) and mode of instruction (exposition, lab work, conceptual models, etc.). In one example, teachers' logs showed similar time devoted to teaching a variety of physical science topics, but student outcomes differed depending on the type of instructional strategies used. Students whose teachers used mainly expository techniques spent over half of their time memorizing facts, while students whose teachers used more lab work spent time collecting and interpreting data, ordering and estimating, and solving routine and novel problems. These studies suggest that any equity metric must go beyond counting minutes, hours, or courses. However, indicators of enrollment in key gatekeeping courses (eighth-grade algebra and high school geometry) and indicators of the *Availability of Advanced Science and Mathematics Courses* are critical. The analysis of indicators across studies also suggests that academic tracking of students can be identified most easily by assessing both the intent to enroll in an *Academic Program* in the eighth grade and actual enrollment in the tenth grade by subgroups of students.

NSF's statewide systemic initiative effort has identified two additional variables, neither of which is easy to quantify (Shields, Marsh, & Adelman, 1997). Like other indicators in the metric, both argue for new methods of assessment. One of these variables is *Instructional Quality*; the other is *Quality of Teacher Professional Development*. Quality of instruction includes multiple ways of engaging students in active exploration and problem-solving, involving students in curriculum decisions, using multiple methods to systematically gather data about student understanding and ability, encouraging students' questions and information sharing, and working in small, cooperative groups.

The other variable, *Quality of Teacher Professional Development*, is an overall indicator of movement toward equity. Challenging curriculum, varied instructional strategies, as well as increased understanding of the backgrounds of diverse subgroups of students, require life-long learning and skill development. Measures of the *Quality of Teacher Professional Development* need to move beyond counting the number of college or continuing education credits accrued to assessment of the quality of outcomes. That is, evidence of changing practices, behaviors, and attitudes is needed—evidence that may be collected through teacher logs, student journals, audio and video tapes, and interviews, as well as by improvement in the retention and achievement of subgroups of students.

Moving toward equity is a stepwise progression, one that will require time to collect data, instigate change, and assess trends. It is anticipated that systems will need at least five years to document substantive improvement in the four areas identified in Figure 2. First, current policies and practices need to be reviewed. Are enrollment and achievement data disaggregated so that the progress of subgroups may be plotted? Are results available by school and course to principals and teachers so that change can be affected? Are there embedded structures (free course selection, attendance taken only in first period, limited scheduling) that limit access or retention?[6] Next, teachers and administrators need to develop and gain community acceptance of an equity plan. Incentives for change need to be identified and implemented. There has always been a reliance on the good will and the nurturing nature of teachers and administrators, but will internal incentives provide sufficient rewards for this very difficult process? External incentives will probably be needed. Public school systems have tried incentives of pay and professional rewards in the past, and most trials have been short live. Identification and implementation of appropriate incentives is a difficult task, complicated by union rules and the public's expectations, but a few successful trials indicate that appropriate incentives can be identified and justly administered. Each system will have to conduct its own review and develop its own equity plan, and each will also want to identify and institute incentives.

The complexity of Figure 2 suggests that a unilateral approach cannot be taken in developing an equity metric; that is, one size does **not** fit all students or systems. Rather, indicators are needed that assess variation in group needs, that take into account the qualitative differences in experience when a subgroup reaches a *critical mass* (the number participating by subgroup is such that the participation is not considered exceptional), and that are sensitive to what particular student groups bring to the classroom. One caution is that indicators may be used in two ways: in one way they may serve to identify gateways for subgroups, but there is the danger that they could serve as gatekeepers to learning opportunities (W. Clune, personal communication, December 6, 1996). Further, any equity metric will need to consider the resources of the system to which it is applied as well as the levels of equity within that system. The first point in assessing equity is to ascertain whether the system is addressing the indicators of differences in access, retention, and achievement.

Measurement and Psychometric Issues—The Tip of the Iceberg

> The difference in achievement between minority and other students is 45 percent less if **selected** (authors' emphasis) home, school, and individual characteristics are equal. This finding suggests that by changing some variables such as high school curriculum and educational activities at home, student learning in mathematics and science could be improved significantly.
>
> NCES, *Understanding Racial-Ethnic Differences*

Clearly, the first and most important caution is that indicators for an equity metric do not address disparities in socioeconomic status, yet poverty in the home as well as attendance at an economically disadvantaged school contribute significantly to differences in science and mathematics achievement. For example, NCES regression analyses indicate that 29 percent of the achievement difference between white and Asian American students (one group) and African American, Hispanic, and Native American students (second group) is related to differences in resources and learning activities in the home. In mathematics, those indicators account for 36 percent of the achievement differences between those two groups (NCES, 1995). The point is that any equity metric only partially measures factors affecting access, retention, and achievement in mathematics and science. However, taken together, applied appropriately, and interpreted with caution and intelligence, a mix of indicators can measure a system's progress toward equitable education for all students.

There are, however, important measurement and psychometric issues to consider. First of all, any assessment system used must provide disaggregated data so that outcome differences between groups can be analyzed and understood. Second, because student outcome data are the crux of the equity metric, appropriate models for collecting and analyzing achievement data are a critical issue. Witte (1996) argues that both norm-referenced and criterion-referenced tests may be used. His rationale is that systemic reform assessments must include tests that meet the following requirements: (1) they can be administered to relatively large populations, and (2) they can be scored relative to a common interval metric that can be converted into an easily understood probability distribution. Tests may include typical norm-referenced ones in which items are designed to measure and compare student mastery of skills and knowledge. In addition, criterion-referenced tests can be used if the test criterion levels allow measurement and comparison of student mastery. For example, when NAEP standard scores are converted to proficiency levels, they meet this requirement; where they provide measurements and comparisons of student skills and knowledge, performance-based assessments and essays also can be used, as long as the scoring produces a metric that can be used for comparison of student mastery.

Witte (1996) suggests that trend data—data that acknowledge the long-term aspect of systemic reform and that provide accurate estimates of achievement growth—may be particularly appropriate. He postulates two models for assessing student achievement in systemic reform: the cohort point-estimate model and the value-added model. The cohort point-estimate model assesses different cohorts of students at the same point in time over several years (e.g., scores on a tenth-grade algebra test), while the value-added model follows one cohort of students over time (e.g., mathematics scores throughout high school). Witte argues that there is a need to measure the value added by the system; that is, the change in achievement the system has produced over time. He recommends multivariate estimates of achievement, using information schools have in their databases, to provide an accurate and rich picture of learning outcomes. Those estimates will provide more meaningful information than can be obtained by reporting simple, mean point-estimates of achievement by different cohorts of students at the same point in time. He maintains also that the value-added model provides more information than a comparison of a group of students to a national population of students on a single administration of a test, as is commonly done now. Following Witte's cautions and recommendations in the collection and interpretation of student achievement data would provide more accurate assessments of student subgroups and of the nature and level of equity in a system.

Boone's (1998) work in assessing a statewide systemic initiative has identified another caution: not all subgroups of students have the same patterns in test completion. He proposes ways in which test construction can guard against instrument bias. For example, Boone found that girls and boys as well as white and African American children have distinct patterns in test taking, with both girls and African American students completing significantly fewer items near the end of the test. He cautions that the design of the test can greatly influence the quality of achievement measures calculated for students and recommends both mixing items by difficulty throughout any test and using statistical analyses that enable one not to count missing answers as incorrect ones. Considering his findings, it is important to separate out responses that are missing because a student "did not get to the item" from ones unanswered because a student "did not know the answer." Interestingly, Boone's analysis suggests another indicator of equity; that is, equal completion rates on multiple-choice tests (in terms of the "did not get to") for subgroups of students.

In addition, sampling techniques may be utilized to provide more comprehensive information at less cost. In general, Witte (1996) warns that "the tradeoffs [in the assessment of systems] are numerous. More reliable and valid test instruments may take more time to administer and score, and hence cost more" (p. 5). Although the indicators, identified through analyses of large databases, suggest elements to assess in a system's progress toward equity, issues of sampling and analysis are complex. These issues will have to be factored into the way that indicators of equity are measured in any system. From this brief review, it is clear that any assessment will only describe the tip of the iceberg.

Clearly, the task is to identify indicators that are directly linked to outcomes (student access, retention, and achievement). The indicators and the measures used will need to vary over time as different aspects of the system are assessed or as different subgroups emerge. In addition, data will need to be collected from multiple sources, such as teacher logs, student portfolios, performance assessments, and direct classroom observations. An equity metric will need to be sensitive enough to show progress being made by different subgroups, to be valid and reliable, to be meaningful and credible to educators and the public, to be understandable to important audiences, to be flexible in terms of situations and levels of education, to have multiple parts, and to have content specificity (i.e., sensitivity to issues in science and mathematics). The goal is to assess when a system is making progress toward equity. Multiple indicators, assessed in valid and reliable ways, provide the first step. Each system will need to formulate equity goals that address the indicators pertinent to its situation and to monitor its progress towards those goals.

Putting It Together—An Equity Metric

Because each system, indeed each school, will need to tailor any equity metric to its own situation and conditions, any example is limited. However, research and analysis of national databases suggest key indicators as well as ones that are applicable for specific subgroups of students. Further, various indicators may be applied most appropriately at different levels of an educational system. With the caveat that one size does **not** fit all, what are key indicators that a system is becoming more equitable? First, in all three of the large databases, there is clear evidence that placement and achievement in eighth-grade mathematics, not science, is key to future success. Specifically, access, retention, and achievement in eighth-grade algebra are key indicators of a student's probability of achieving a high quality education in mathematics and science. Second, although not easily quantified, the quality of the content and instruction of science and mathematics courses is critical (Hiebert et al., 1997). The criticism that our curriculum is a "mile wide and an inch deep" hits at the core of quality (Schmidt, McKnight, & Raizen, 1997). Third, a clear indication of progress is provided by data from achievement tests that show narrowing of gaps concomitant with rising achievement levels of all students (Witte, 1996). This type of achievement data has been selected as a key indicator because the results are easily understood by important audiences (Kahle, 1997). Fourth, evidence that teaching practices are changing in ways that involve students actively in learning is important, because active involvement increases the probability of both understanding and retention. Although it is tempting to continue to identify key indicators, these four will indicate movement toward equity and provide salient guideposts along the way.

Another approach is to look for indicators that are key for specific subgroups; that is, which indicators best meet a system's priorities? In a rural system where all children have similar ethnic/racial backgrounds and all speak English in their homes, movement toward equity may involve removing differences between girls and boys. What are key indicators of gender equity? First, the documented decline in girls', compared to boys', attitudes about and interest in science suggests that a key indicator is sustained positive attitudes and interest levels as girls proceed from fourth grade (where they are as positive about and as interested in science as boys are) through high school.[7] Second, evidence of instruction that involves cooperative learning groups and activities related to everyday life and evidence of assessments that include writing and explanation suggest that instruction is meeting the interests and needs of girls.[8] Third, progress is suggested by indications that girls' out-of-school science and mathematics experiences are similar in frequency and type to those of boys (Kahle & Lakes, 1983). Fourth, equal enrollments of boys and girls in high school physics indicates that the system is becoming more equitable.

Different indicators might be the focus of assessment in an urban system whose identifiable subgroups are African American students and white students. Key indicators that such a system is moving toward meeting the needs of the African American girls and boys who are underrepresented in terms of enrollment and achievement in science and mathematics courses are increased enrollments in preschool programs, proportional enrollment and achievement in eighth-grade algebra, availability of science and mathematics courses that meet the national science and mathematics standards, increased representation of African American students in academic programs in high school, a decrease in the acceptance or use of behaviors that detract from learning, and proportional enrollment in calculus.

Similarly, one may identify indicators that are key to a school's progress toward equity as well as those that suggest movement of a whole system. Each school will want to monitor changes in course enrollments and the achievement levels of subgroups as well as the quality of the curriculum and instruction provided. Student enrollment in the academic program is best monitored at the school level, as are teacher expectations and behaviors and their effect on subgroups of students. The system, however, is positioned to assess any patterns in teacher or student mobility and their effect on equitable education and to implement policies that either decrease mobility or compensate for it. In addition, positive changes in graduation rates, graduation requirements, and standardized test scores that are monitored centrally can provide evidence of a more equitable system. Further, it is the responsibility of the system to assess the effectiveness of teacher professional development in moving the system toward equity and to change or alter those experiences as needed.

Each system is unique and each equity plan will involve different components. An example of an equity metric suggests how one can work. Central City School Corporation (CCSC) enrolls 70 percent African American and 30 percent white (largely Appalachian) students. The district's elementary, middle, and high schools are divided among magnet schools, neighborhood schools, and neighborhood schools with magnet programs; this complex mix is the result of court-ordered desegregation guidelines and quotas. Before initiating its systemic reform and equity plan, data from CCSC's evaluation division indicated extensive tracking of middle and high school students into basic, general, and academic courses in mathematics and science. In addition, over half of the African American students failed ninth-grade algebra and biology. When the state initiated proficiency examinations, higher proportions of African American than white students failed them. Further, over half of the students who entered high school dropped out prior to graduation, and the rate was higher for African Americans. However, the district had a strong program in advanced placement courses and equal numbers of African American and white graduates entered college. Because data were not disaggregated by race and gender, issues of gender equity had not been addressed. In the area of professional development, district teachers had an option of attending courses of their choice at several area universities to move up the salary schedule and professional ladder.

With the advent of both local and national reform, CCSC charted a plan to move toward meeting the needs of all children, equalizing opportunities to learn across courses and schools by providing equitable education. Although district administrators and teachers realized that many aspects of the system would need to be evaluated, they chose to begin with opportunities to learn and achievement. First, a comprehensive assessment plan was created so that baseline as well as trend data were available to guide the reform. Initially, CCSC chose the following measures to assess academic progress in science and mathematics by race/ethnicity and gender.

4th grade	Stanford 9 Test of Achievement State Proficiency Tests in Mathematics and Science Minutes/Day of Instruction in Science and Mathematics Student and Teacher Mobility
8th grade	Stanford 9 Test of Achievement Instructional Assessment Tests (MetriTech Co.) State Proficiency Tests in Mathematics and Science Enrollment in Mathematics by Course Selection of Academic Program Student and Teacher Instructional Practice Surveys—Horizon Research, Inc. [HRI], Local Systemic Change Initiatives <http://www.horizon-research.com/LSC/default.htm>
10th grade	Passing Rates in Algebra and Biology Enrollment in Geometry Retention in Academic Program Student Mobility (Including Dropout Rates) Teacher Mobility
12th grade	State Proficiency Tests in Mathematics and Science Advanced Placement Scores SAT and ACT Scores Number of Science and Mathematics Courses Completed Graduation Rates College Entrance Rates

All data were analyzed by both race and gender to identify any differences among subgroups, and individual school data were returned to the principals and teachers for discussion and action. As the reform progressed, CCSC (with its union's support) requested that schools set equity goals and provided incentives for reaching the goals. Principals' raises were linked to improvement, as were school-based bonuses.

CCSC instituted curricular reforms (both content and instruction) and developed mechanisms for monitoring progress. All remedial and general mathematics and science courses were identified (a variety of course titles had been in place) and replaced by academic courses, and reviews of student transcripts provided progress data. Research-validated inquiry-based curricula were identified, and professional development was provided for school-based teams of teachers.[9] Teachers kept logs and the district surveyed a random sample of teacher logs and student portfolios to assess changes in teaching practice and in the implemented curriculum.

To address the critical issue of unacceptable failure rates in biology and algebra as well as high school dropout rates, the district collected data on student and teacher mobility and began to allow students to complete the school year in the same school, regardless of geographic boundaries. Incentives were provided to encourage teachers to continue at the same school for a period of years. Elementary and middle schools were reorganized into multilevel teams so that teachers and students had the opportunity of becoming learning communities. Attitudinal data (e.g., the Fennema-Sherman scale, 1976), behavioral data (numbers of in- or out-of-school suspensions), and attendance data (by specific course) were collected to indicate progress or problems by subgroups. Further, the system instituted summer programs between the eighth and ninth-grades for eighth-grade students who were at risk of failing algebra and/or biology. The failure rates dropped precipitously, indicating movement toward equity and the need for similar bridge programs throughout high school.

As the reform matured, analyses of teaching practice and achievement data continued to identify leverage points in the system. In addition, it was possible to show the positive effect of a critical mass of minority students in a calculus class on both their achievement and future educational goals and to change boundaries and scheduling to ensure a critical mass in other indicator courses.

As the district's white population became increasingly Appalachian, appropriate indicators were added to the equity plan. For example, attendance in preschool, students' beliefs about the usefulness of mathematics and science, and course selection patterns were monitored for indications of inequity.

CCSC's experience may be useful as an example, but each system has different resources and different subgroups, and each is at a different place on the pathway to equitable education. Components of any equity metric, no matter how well conceived, will change as new subgroups emerge and as educational systems evolve—the indicators and the measures needed in the early parts of the reform may not be appropriate for its later stages. CCSC, for example, found that past measures of student achievement did not reflect the content of its new inquiry-based curricula. CCSC valued student achievement at the fourth, eighth, tenth, and twelfth grades as indicators of progress and problems, but it needed new achievement measures, such as tests composed of public-release NAEP or TIMSS items or new performance-based assessments.

In biology, systemic means affecting the whole system (nervous, digestive, etc.), and each system has self-correcting feedback mechanisms. In education, systemic reform also refers to the whole system, affecting all parts. An equity metric may be used by administrators and teachers to provide continuous feedback during systemic reform, informing and changing components as needed, addressing and correcting inequities, and evolving and adapting indicators and measures. It is not the one, or only, solution, but it may allow reformers to assess progress and to alleviate problems in providing equitable education in science and mathematics for all students.

Endnotes

[1]. For a complete description of these studies, see NELS:88 (Ingels, Abraham, Karr, Spencer, & Fraukel, 1989), High School and Beyond (Peng, Fetters, & Kolstad, 1981), and TIMSS (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996).

[2]. Because the terminology of the U.S. Census to identify specific ethnic/racial groups is used in most large national studies, it is used here. Although the concerns about using one or more of the Census' terms are understood and valid, those terms are applied across all references (except in quotations) in this paper.

[3]. The recent court rulings that reinforce equal numbers of intercollegiate athletics programs for males and females provide a situation in which an identifiable subgroup (male athletes) may experience discrimination in opportunities (elimination of some sport teams). Given the historical disadvantage of women and the legal requirement to provide equal numbers of intercollegiate teams, this type of discrimination is based on ethically sufficient reasons.

[4]. In standards-based educational reform, local performance standards should ensure that all students achieve mathematics and scientific literacy as defined by the standards of the National Council of Teachers of Mathematics (1989) and the National Research Council (1996).

[5]. According to a Census survey of 55,000 households in 1995, the proportion of 25- to 29-year-old Hispanics and Asian Americans with high school diplomas declined between 1992 and 1994. The survey also found that comparable percentages of 25- to 29-year-old African Americans (86.5%) and whites (87.4%) held high school diplomas in 1995 (Holmes, 1996). However, the most recent *Condition of Education* (Smith, 1997) reports that differences exist in the high school graduation rates of African Americans and Whites.

[6]. In small middle and senior high schools, the scheduling of one section of band, orchestra, or second-year French, for example, effectively tracks students into certain sections of other courses. Similarly, there may be hidden tracking in schools that have team-based instruction. Although all teams have heterogeneous groups of children, and thereby meet the district's mandate to eliminate academic grouping or tracking, individual teams may group children by ability.

[7]. There is a less dramatic decline in girls' interest in and positive attitudes about mathematics, so attitudes about science have been selected as the key indicator (Kahle, 1996).

8. Gender equity research indicates that girls prefer to learn in cooperative groups and to have science instruction related to real life experiences. Further, there is evidence that girls perform better on written, compared to multiple-choice, assessments (Fennema, 1990; Kahle, 1996).

[9]. A sample of the curricula that meet the criteria include: *Foundational Approaches to Science Teaching* (FAST), *Full Option Science System* (FOSS), the BSCS programs, the *Connected Mathematics Project* (CMP), *Algebra Project*, *Physics by Inquiry*, as well as the professional development program *Cognitively Guided Instruction* (CGI).

---

[1]. For a complete description of these studies, see NELS:88 (Ingels, Abraham, Karr, Spencer, & Fraukel, 1989), High School and Beyond (Peng, Fetters, & Kolstad, 1981), and TIMSS (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996).

[2]. Because the terminology of the U.S. Census to identify specific ethnic/racial groups is used in most large national studies, it is used here. Although the concerns about using one or more of the Census' terms are understood and valid, those terms are applied across all references (except in quotations) in this paper.

[3]. The recent court rulings that reinforce equal numbers of intercollegiate athletics programs for males and females provide a situation in which an identifiable subgroup (male athletes) may experience discrimination in opportunities (elimination of some sport teams). Given the historical disadvantage of women and the legal requirement to provide equal numbers of intercollegiate teams, this type of discrimination is based on ethically sufficient reasons.

[4]. In standards-based educational reform, local performance standards should ensure that all students achieve mathematics and scientific literacy as defined by the standards of the National Council of Teachers of Mathematics (1989) and the National Research Council (1996).

[5]. According to a Census survey of 55,000 households in 1995, the proportion of 25- to 29-year-old Hispanics and Asian Americans with high school diplomas declined between 1992 and 1994. The survey also found that comparable percentages of 25- to 29-year-old African Americans (86.5%) and whites (87.4%) held high school

diplomas in 1995 (Holmes, 1996). However, the most recent *Condition of Education* (Smith, 1997) reports that differences exist in the high school graduation rates of African Americans and Whites.

[6]. In small middle and senior high schools, the scheduling of one section of band, orchestra, or second-year French, for example, effectively tracks students into certain sections of other courses. Similarly, there may be hidden tracking in schools that have team-based instruction. Although all teams have heterogeneous groups of children, and thereby meet the district's mandate to eliminate academic grouping or tracking, individual teams may group children by ability.

[7]. There is a less dramatic decline in girls' interest in and positive attitudes about mathematics, so attitudes about science have been selected as the key indicator (Kahle, 1996).

8. Gender equity research indicates that girls prefer to learn in cooperative groups and to have science instruction related to real life experiences. Further, there is evidence that girls perform better on written, compared to multiple-choice, assessments (Fennema, 1990; Kahle, 1996).

[9]. A sample of the curricula that meet the criteria include: *Foundational Approaches to Science Teaching* (FAST), *Full Option Science System* (FOSS), the BSCS programs, the *Connected Mathematics Project* (CMP), *Algebra Project*, *Physics by Inquiry*, as well as the professional development program *Cognitively Guided Instruction* (CGI).