DOCUMENT RESUME

ED 363 617 TM 020 455

AUTHOR Stewart, Donald M.; Everson, Howard T.
TITLE Toward National Standards and Testing: The Educational Equity Imperative. Draft.
PUB DATE Mar 93
NOTE 23p.; Paper presented at the Ford Foundation National Symposium on Equity and Educational Testing and Assessment (Washington, DC, March 11-12, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Standards; Access to Education; College Entrance Examinations; *Educational Assessment; Educationally Disadvantaged; Elementary Secondary Education; *Equal Education; *Evaluation Methods; Excellence in Education; Higher Education; Mathematics Achievement; Minority Group Children; *National Competency Tests; Racial Differences; Track System (Education)
IDENTIFIERS College Entrance Examination Board; EQUITY 2000; *Evaluation Standards; *National Standards; Performance Based Evaluation; Project PACESETTER

ABSTRACT

A brief sketch is presented of the assessment strategies of the College Board in two unique programs, EQUITY 2000 and Pacesetter, both works in progress that attempt to meet the challenges of the new era of equity and excellence in education. EQUITY 2000, launched in 1990, is a multi-faceted approach to closing the gap between college-going rates of minority and non-minority students currently being implemented in six urban school districts. It is an intervention that emphasizes all aspects of a child's education, focusing on mathematics at the middle school and high school levels, and attempting to eliminate tracking in mathematics courses as a way to encourage other significant curriculum reform. Current EQUITY 2000 plans call for using a standardized test to measure achievement. In Pacesetter, the College Board will develop both formative and summative performance-based assessments for particular courses. Pacesetter also aims to eliminate tracking in school systems. Pacesetter courses will include mathematics, English, world history, science, and Spanish. Professional development will be an essential part of the Pacesetter process. Both initiatives are designed to help create an educational system in which the vast majority of students are exposed to quality education that only the top third have been exposed to up to now. (Contains 23 references.) (SLD)

DRAFT

Not for quotation or attribution without permission of the authors.

TOWARD NATIONAL STANDARDS AND TESTING:

THE EDUCATIONAL EQUITY IMPERATIVE

Donald M. Stewart

and

Howard T. Everson

The College Board

New York, NY

A paper presented at the Ford Foundation sponsored *National Symposium on Equity and Educational Testing and Assessment,* Washington, DC, March 11-12, 1993.

The decade of the 1980s brought a renewed focus on educational reform. With the publication of *A Nation at Risk* in 1983, and a subsequent wave of reports documenting the lamentable state of education in America, greater attention has been directed towards ways to better educate our students. The consensus emerging from more than a decade of debate is clear: America's schools have to do a better job. In response to this challenge, business representatives, political leaders, and the educational community itself, are calling for a broader use of educational tests and assessments to measure the outcomes of schooling and gauge the educational system's successes.

Yet, as we approach the 21st century we find ourselves in the midst of a revolution in educational testing and assessment. Fundamental changes, both in the role of educational assessment and how those assessments are conducted, are upon us (Sizer, 1985; Stiggins, 1991). More to the point, the field of educational assessment, many believe, is undergoing a distinct paradigm shift (Kuhn, 1970), moving away from a long-standing reliance on norm-referenced, multiple-choice type tests while, at the same time, embracing aggressively an array of alternative performance assessments thought to be more "ecologically" valid.

More important for testing and assessment, what we know about learning and instruction is changing (Herman, 1991; Wittrock, 1991). Within the last two decades or so a steady flow of educational and psychological research has focused on the cognitive processes of students and teachers. This new knowledge has reshaped our understanding of how students learn and, more to the point, how tests should be designed to improve instruction. These new insights into the teaching and learning process have generated changes in the educational outcomes that society values. No longer is there an emphasis on having our children demonstrate what they know and can do by relying on rote memorization of facts or by displaying discrete, and often mechanical, skills in overly narrow academic domains. Today, the value is on higher order thinking skills, problem solving, and the application of knowledge and skills to settings that transcend the

classroom. The upshot is a growing demand for assessments based on demonstrated performance of these newly valued educational outcomes (Sizer, 1985). Performance based assessments, (e.g., portfolios, complex projects, experiments, and simulations) are rapidly becoming the method of choice for many educators, policy makers, and parent groups (Frechtling, 1991). Indeed, there now appears to be what Dennie Palmer Wolf (1990) has referred to as a "romance of alternative assessment".

Performance assessments, when well conceived and well developed, have much to offer. They are often viewed as more "authentic", looking more like the kinds of instructional tasks we want students to be able to perform. They are complex, often more lifelike (Frechtling, 1991). Proponents of performance assessments also argue that because these approaches have a fair amount of educational currency this approach to assessment will have a greater positive influence on instruction.

They are not without problems, however. As Frechtling (1991), and more recently Koretz (1993) and Worthen (1993) and others have argued, performance assessments are time consuming, costly, and are often subjective and difficult to score. On balance, we can expect hard questions to be asked about this new approach to educational testing and assessment.

Along with the pressure brought on by the new zeitgeist in assessment we can, for example, expect an increase public scrutiny, as well. As we move on the state and district level toward using complex performance assessments in large scale educational testing programs, calls for public accountability to ensure fairness, equity and integrity in educational testing can be expected to mount, particularly as the results of some of the early attempts at implementing performance assessments on a large scale are studied and understood (Koretz, 1992; Maeroff, 1991). Thus, the challenge in this changing era of educational assessment is to develop a new generation of tests and assessments that meet the dual goals of fostering higher standards and promoting educational equity for all students. In this paper we present a brief sketch of the *College Board's* assessment

strategies in two unique programs *EQUITY 2000* and *Pacesetter* , both "works in progress" which attempt to meet the challenges of this new era, particularly the challenge of equity and excellence. In our view, the success of *EQUITY 2000* and *Pacesetter* will not only require creativity in the design of new modes of assessment, it will demand that the equity imperative be central to the design of all educational tests and assessments. By contrasting the differences in the tests and performance assessment we are planning for these two major educational initiatives, we hope to highlight the implications for equity in large-scale assessments more generally.

Educational Equity: An Imperative

Promoting equity in academe has become a crucial goal for educators in America today. The civil rights movement of the 1960s, and the resulting programs and policies aimed at ensuring the educational opportunities for Blacks, women, and other minority groups, were only partially successful. If America is to expand its economic competitiveness and provide for the general education of each of its citizens, then it must find a way to diversify its educational institutions. In particular, given the growth in the demand for educational assessment, we must strive for and achieve equity in the arena of educational testing, an area so central to educational reform.

Equity, however, is more than a philosophy. A serious and ongoing commitment to fostering equity in American education goes beyond changing values and attitudes to include changing policies, programs and practices that affect learning and achievement. Equity becomes actions and behaviors that help create educational outcomes with no differences based on race, sex, or economic status. Within the framework of educational testing and assessment, the equity imperative translates into a number of clearly articulated principles and practices, including, for example, the belief that all those being assessed have the benefit of prior exposure to instructions, materials, and equipment that adequately prepare them for the experience. Test developers, too, must give careful attention to the

content, language and form of the tests so as not to disadvantage some students because of their membership in any ethnic group, sex, social class or physical handicaps. More important, the results of these assessments must be equally valid and reliable for members of each ethnic or racial group, sex or social class.

As those with the responsibility for developing assessments and monitoring movement toward national standards, we need to issue guidelines prescribing the appropriate uses of test scores and other results. We ought to provide information about interventions and instructional materials needed by students to achieve the high standards represented in the assessments

The *College Board* is well known today for its assessment programs. Historically, the *College Board's* push for equity can be traced to the development of the *SAT* and the insistence that access to college be based in part on a common test (Stewart, in press), an equitable yardstick. Clearly, much has been gained. Today, the *SAT*, the *PSAT/NMSQT* and the *Advanced Placement (AP) Program* are taken by millions of students each year, and the impact of these assessment programs is felt throughout the educational system.

Ensuring that these assessment instruments are free of bias is central to the *College Board's* commitment to educational equity; it is at the very heart of the validity and utility of our tests. Evidence of the extent of our commitment can be found in the volume of psychometric research devoted to maintaining bias-free tests is widely available (Berk, 1982; Cole, 1981). The *College Board* and its partner the *Educational Testing Service* are recognized leaders in the field of psychometric research, particularly research focusing on test bias and test validity (Wainer and Holland, 1992). Like the larger paradigm shift we are witnessing in educational assessment, research into the phenomenon of test or item bias is shifting its focus as well. Many researchers and measurement experts are moving from an emphasis on the test or the test items as the genesis of differential performance to a view that questions the cognitive factors, background experiences (Muthen, Chih-Fen, and Burstein, 1991; Skaggs and Lossitz, 1992) and opportunities to learn (Anderson, 1985)

that are reflected in the various assessment instruments. This changing notion of equity, the idea that subgroup differences seen in test scores and performances are conditioned on a host of cognitive factors, like instructional opportunities or cognitive learning styles (see Johnson, in press), has profound implications not only for equity in our educational assessments, but also for the fit between assessment and instruction for all students. It is this "goodness of fit"--the articulation of instructional objectives and assessment methods--that *EQUITY 2000* and *Pacesetter* attempt to address.

Clearly, equity in educational testing and assessment is more than a philosophy, it is central to the mission of the *College Board.* To underscore that commitment, the Trustees of the *College Board*--elected representatives of more than 2,800 member schools, colleges, universities and educational systems--recently approved revisions in the organization's mission statement to address the challenges of this new era in educational assessment by advocating explicitly..."universal access to high standards of learning, equity of opportunity,... so that every student is prepared for success in college and work." This new mission represents a clear mandate for change in educational assessment and testing. It establishes the equity imperative in educational testing and assessment.

Vision and Innovation: Ingredients for Reform

Over the years, the *College Board* has been a significant contributor to the reform of education and has often been a catalyst for change. Between 1980 and 1990, for example, the *College Board* sponsored the *Educational EQuality Project* (*EQ*), a national public service program to strengthen the academic quality of high school education through subject area standards and teacher development. Since 1983 over 450,000 copies of EQ's first major contribution to school reform *Academic Preparation for College* have been distributed to school districts and states as a guide to curriculum reform; literally hundreds of *EQ* workshops have been given across the country.

Schools and colleges, for some time now, have recognized that some students are capable of accomplishing college-level studies while in high school. The *College Board's Advanced Placement* (*AP*) program is a means by which secondary schools and colleges can provide such educational experiences. *AP* provides descriptions of and examinations in college level courses to secondary school students, and the results of the examinations to colleges. Teacher institutes are offered regularly throughout the country to help new and veteran *AP* teachers prepare their students for guides and examinations in college subjects. *AP*, through committees of school and college faculty, also develops course descriptions, curricular outlines and teaching guides. The *AP* exams, which require students to perform a variety of challenging tasks and solve a range of complex problems, are a model for many advances in performance-based assessment. *AP*, in fact, is one of the fastest growing educational programs in the nation.

In a continuing effort to facilitate the transition from high school to college, the *College Board* in the Spring of 1994 will introduce a new *SAT* . Like its well known predecessors, this new assessment will drive change in educational testing by introducing many innovations in assessment and measurement technology, including critical reading tasks, allowing student produced responses, as well as permitting calculator use in the mathematics sections of the test. The new *SAT*'s high level reasoning tests in both the verbal and mathematical domains, for example, reflect the need to measure higher order cognitive abilities important for success in college. The recommended use of calculators, no doubt, will foster changes in mathematics instruction and bring about an increase in the opportunities for students to use micro-chip technology in the service of learning. The performance-based assessments in writing, mathematics, and other areas signal students, teachers and parents of the need to acquire higher order skills and to practice complex performances. These changes to the *SAT* also ensure the test better reflects current curricular and instructional practices in our nation's secondary schools; these changes bring us one step closer to integrating measures of reasoning and achievement.

Innovation in educational assessment without a vision for excellence and a focus on purpose will achieve little and, at best, foster only modest reform. In addition to current programs that set a high, external standard for student accomplishment, the *College Board's* vision focuses on reform through a dynamic change process, a "push-pull" strategy wherein educational assessments both push and pull the system toward higher standards and outcomes for all students. This approach, with *EQUITY 2000* as the "push" and *Pacesetter* as the "pull", will address the challenge of ensuring equity in the outcomes measured by our educational tests and assessments.

The Push-Pull of Educational Reform

Over the years, there have been many signs of concern about, and desire for, truly national educational standards. Having high goals, however, will not, in and of itself, produce the desired higher achievement for all students, especially given the enormous variety of economic, ethnic, linguistic and family backgrounds in our nation. What we need now are implementation strategies for achieving these high standards, while at the same time ensuring educational equity.

The *College Board* is working in two directions that help with these very difficult implementation issues. First, if all students are to reach these new high standards at graduation from high school, efforts for higher standards of preparation in earlier years--middle school and early high school--will be required. Opportunities to learn must be available. The *College Board* is concerned about the extent to which the secondary school mathematics curriculum, i.e., algebra, geometry, and pre calculus, serves as a filtering device in the nation's high schools. It is clear that those students who, early on, take courses in algebra and geometry have the option not only of pursuing courses in higher mathematics, but also do well on standardized tests and other forms of assessment. As a result they gain access to other academically rigorous fields of study in college preparatory courses across the curriculum which are closed to students who do not take those key math

courses. These foundation courses, we believe, set the stage for sustainable learning and achievement for all students. Thus, armed with this knowledge of the effects of course taking and the desire to transform the mathematics filter into a pump (Treisman, 1985)--a force that propels students toward higher standards--we started a project known as *EQUITY 2000* to help schools get all students into pre-algebra in the middle school and into algebra and geometry in high school. *EQUITY 2000* is the "push".

A second, and complementary, effort of the *College Board* is directed at helping schools implement higher standards by providing new course syllabi for high school courses in key subjects, which along with appropriate training, help teachers learn to teach to higher standards, as well. We call these *Pacesetter* courses, and we are designing them with the help of disciplinary associations so that schools wishing to raise their standards will have course materials, teacher preparation assistance, and assessments to do so. The *Pacesetter* courses in mathematics, English, science, world history, and Spanish at the upper secondary school level are the magnetic force--the "pull"--at the top of the secondary school curriculum. But first, more needs to be said about the "push", *EQUITY 2000* .

EQUITY 2000

In 1990, the *College Board* launched *EQUITY 2000*, a multi-faceted approach to closing the gap between the college-going rates of minority and non-minority students. *EQUITY 2000*, currently being implemented in six urban school districts across the country, is an intervention program that emphasizes all aspects of a child's education: curriculum, quality of instruction, environment and equal access to all learning opportunities. *EQUITY 2000* focuses on mathematics at the middle school and high school levels, and specifically aims to eliminate tracking in mathematics courses as a way to encourage other significant curriculum reform.

Working with major foundations, and a consortium of teachers and researchers, we have created *EQUITY 2000,* as a partnership with urban school districts. This is a national

$30 million program that requires that all middle and high school students learn algebra and geometry, and receive strong pre-collegiate counseling designed to ensure awareness of the opportunities for successful admission to, and achievement in, college. Fundamental to *EQUITY 2000* is the philosophy that by offering all students high standards and opportunities to achieve them, the gap between majority and minority student achievement and success will be closed. The expectation is that all students can learn and will achieve given adequate resources, encouragement and support. The aim is to end tracking so that all students take academically rigorous courses across the curriculum. Algebra and geometry are the gatekeepers. *EQUITY 2000* will transform those courses into gateways.

Involving a significant component of professional training for teachers and guidance counselors, as well as a sustained effort to involve school superintendents, principals, school board members and community leaders, *EQUITY 2000* is being piloted in six sites around the nation. Starting slightly over two years ago, these sites are, Fort Worth, Texas; San Jose, California; Prince George's County, Maryland; Providence, Rhode Island; Milwaukee, Wisconsin and Nashville, Tennessee. Two things are now very clear. The first is that motivation and the expectation of success in learning can be significantly raised among both students and teachers. The second, is that the desire to undo the pernicious practice of tracking has begun to ripple through the schools and systems involved with *EQUITY 2000*.

Current plans for student assessment in *EQUITY 2000* call for using a standardized, norm-referenced test--the *PSAT*, in all likelihood--to measure mathematics achievement. Thus, like many traditional educational programs and interventions, student achievement in *EQUITY 2000* will be measured through the use of an educational test that is largely external to the instructional process. The *PSAT* mathematics test, which is a derivative of the *SAT*, presents students with a number of multiple-choice items measuring quantitative reasoning. In addition to standardized test scores, students' attitudes and course taking behaviors will also be measured. Moreover, the educational achievements of *EQUITY*

2000 students will also be followed over time. Future performance on the *SAT* and in college-level courses will also be monitored.

The choice of a standardized, multiple-choice type test is deliberate. The psychometric properties of the *PSAT* and *SAT* are well known. Also well known are the performance differences between various subgroups on these types of measures. By gathering performance data in this manner, however, we have the unique opportunity to examine the students' ability estimates in mathematics when exposure to instruction and the opportunities to learn mathematics has been assured. Subsequent analyses, which will employ psychometric models that are sensitive to instructional opportunities and cognitive background factors (see Muthen, et al. 1991, for example), will allow us to examine the emerging hypothesis that subgroup performance differences are conditioned by prior learning experiences. The notion that *all* students can learn and achieve high standards in mathematics will be tested in *EQUITY 2000* using performance data gathered by traditional forms of educational assessment. *Pacesetter*, on the other hand, takes an entirely different tack when it comes to the link between instruction and assessment.

Pacesetter

Although similar to *EQUITY 2000* in its challenge to foster high standards, in *Pacesetter* we will develop performance-based assessments that are both formative (i.e., embedded in instructional practice) and summative (i.e., end-of-course assessments that permit comparability across sites and among students). As we noted earlier, contrasting the educational testing and assessment approaches in *EQUITY 2000* with the approaches used in *Pacesetter* highlights the implications for equity in large-scale assessments more generally. Linking the *EQUITY 2000* initiative with *Pacesetter* , we believe, provides the intellectual energy for a powerful engine of change, and furthers our efforts in the direction of sustainable learning and equitable assessment.

The goal of *Pacesetter* is also to undo the pernicious practice of tracking in school systems. Like *EQUITY 2000* , this should take place not just by declaring high achievement standards for all students, but also by realizing them in practice. *Pacesetter* initiatives take as their starting point the development of detailed substantive frameworks each of which specifies the structure and content of what should be taught and learned in a key course of study. Beyond high standards for all students and course frameworks which translate those standards into actual syllabi, *Pacesetter* will offer professional development activities so that teachers will be able effectively to instruct all students; activities in the classroom and imbedded assessments that allow students to translate concepts into hands-on situations, and teachers to evaluate students' skills and understanding, thereby informing instruction; there will be end-of-course assessments to measure student attainment of course objectives.

The specific *Pacesetter* courses will include mathematics, English, world history, science, and Spanish. And each of these *Pacesetter* courses is being developed through task forces in collaboration with the major national discipline associations -- the *National Council of Teachers of Mathematics* and the *Mathematical Association of America*; the *National Council of Teachers of English*; the *American Council of Learned Societies* and the *National Council for the Social Sciences* for the *Pacesetter* course in world history; the *National Science Teachers Association*, and the *American Council on the Teaching of Foreign Languages* for Spanish. The job of these task forces is to identify the information and skills students must know, the methods needed to help students learn them, and the ways to assess achievement that integrate the daily work of students with end-of-course examinations. The task forces are made up of experienced educators in each discipline and include high school teachers, district superintendents, curriculum coordinators and university faculty. In addition to these four groups, the *Pacesetter* course frameworks will also be reviewed by appropriate standing curricular bodies within the respective disciplinary organizations.

These key core high school courses come with a syllabus and assessment for each reflecting high standards which schools could adopt to both raise their own standards and certify student learning. Grades and assessment scores on a set of such courses could be used to demonstrate the levels at which students are achieving for accountability purposes, and, by being grounded in the courses, would help the schools to accept responsibility for self-improvement and to know how to go about doing precisely that. What will the courses be like?

Pacesetter in mathematics is a fourth year high school course emphasizing applications of a range of math functions in the context of real problems. This course, which would normally be taken in the 12th grade but also taken in the 10th and 11th grades by students whose work in the subject is accelerated, focuses on what happens when we confront complex data sets with the need to understand patterns, and to continue research, or to come to conclusions. Teachers put together rich, real-world "case studies" in which linear, exponential and logarithmic functions can be applied to problems in fields like industrial design, inventions, economics and demographics.

The fourth year English course will combine literature and communications and include student portfolios, letters, diaries, speeches and essays. Students will be responsible for both classical and contemporary texts to see how they have understood and understand major concerns of human life. Viewed in this way, to be literate means not just the ability to decipher the literal meaning, but rather to understand and interpret the full range of texts, whether novels, plays, speeches, motion pictures or official documents. In independent reading groups students might pair works like *Othello* and Toni Morrison's *Beloved* looking for commonalties and differences across time.

The science *Pacesetter* course will integrate the physical and biological sciences and introduce concepts through the "big" ideas like evolution and energy. Also a fourth year or "capstone" course, it will bring together problems and theories from several branches of science, like ecology and chemistry, and present students with real-world situations.

The courses in Spanish and world history are conceived of as "cornerstone" courses to give students an understanding of the skills and knowledge they need at the mid-point in high school. The *Pacesetter* in World History cuts across time, civilizations, and geography, and covers topics such as historical change, comparisons of societies and interconnections.

Common to all courses is a focus on a wide understanding of what knowledge is. All told, rather than the passing on of conventional knowledge, the intention of *Pacesetter* is to tilt the curriculum toward application and modeling -- in sum, towards the real world.

Seven school districts, reflecting geographic and demographic diversity, have been selected as *Pacesetter* mathematics pilot sites for the 1993-94 school year: San Diego, California; Broward County, Florida; Prince George's County, Maryland; Battle Creek, Michigan; Charlottes-Mecklenburg, North Carolina; Irving, Texas; and Rutland, Vermont. As part of the effort to demonstrate a "push-pull" effect on curriculum, at least two upper level mathematics teachers from the six *EQUITY 2000* sites (including Prince George's County) will participate in institute training

These sites were chosen with careful attention to localities that already have a broad cross-section of educational reform activities among which *Pacesetter* can take its place. We were also seeking states in which a variety of efforts such as curricular and assessment reform, development of curricular frameworks, textbook adoptions etc. are promoting systemic reform. In addition, we wanted districts and schools representing a broad cross section of different racial and ethnic groups. And finally, we wanted strong geographical diversity, including not only different sections of the country, but urban, suburban, rural districts as well.

In the spirit of systemic educational reform, a basic assumption is that *Pacesetter* courses should have the greatest possible potential for exerting a dynamic influence on the high school curriculum as a whole. Therefore, for the most part, *Pacesetter* courses will be "capstone" learning experiences of the high school curriculum, because these provide an

end point for orienting student preparation in previous years of study; or they may be "cornerstone" courses that suggest the kinds of knowledge and skills students should have by the mid-way point in their high school careers.

All of us interested in educational reform are acutely aware of the distinction between assessment for instructional as opposed to accountability purposes. We mentioned previously that *Pacesetter* draws on two kinds of assessment: classroom (formative) and end-of-course (summative). Assessment of both kinds will include performance-based tasks, essays, projects, case histories, portfolios and multiple choice questions. The assessments embedded in instruction will be used for a variety of purposes -- to help students and teachers evaluate the students' progress and to plan future instruction; to evaluate activities that occur over time, such as projects and portfolios; and to strengthen the teachers' role as an effective and supportive class room facilitator and mentor.

As we noted earlier, some of the current forms of school administered standardized achievement tests focus on the recall of facts and formulas rather than on reasoning, problem solving, communication and connections. By developing instructional units within which assessment opportunities are embedded, *Pacesetter* will alter the relationship between testing and teaching in an effort to improve learning, and to help students become self-evaluative, reflective thinkers. This instructional assessment will provide students with interesting tasks and real data, drawn from a wide range of human activities, to use as the means of developing their understanding, both of the concepts and the applications which make those concepts important to all people in our society.

By contrast with the in-course (i.e., formative) assessments, the *Pacesetter* culminating (or summative) assessments provide a key input to an overall certification that student learning has met the high standards of the *Pacesetter* course. The overall purpose of these end-of-course assessments is to evaluate students in terms of public standards, validate student accomplishments, evaluate both classes and schools in terms of their having achieved standards, provide information to school districts on progress in instructional

approaches, and finally, provide employers and colleges with a range of information about students that may be useful for decision-making purposes.

More specifically, we envision the *Pacesetter* culminating assessments as two to three hours in length and including the following (or more) types of activities:

- a complex task for which students prepare themselves in advance to demonstrate their abilities to apply concepts and skills to real-world problems;
- an extended task which includes a number of short-answer questions, problems, or sub-tasks to draw the student toward a larger task; and
- an integrative task which requires the student to integrate ideas and concepts learned in the course and to provide a brief reflection on the integrating process.

Thus, this culminating assessment would call for 1) work such as application or evaluation in the context of a prepared task; 2) analysis, problem solving, or problem construction in the context of a guided task; and 3) integrating and reflecting on a task that requires a student to draw freely on course content and his or her own strengths and interests. This assessment might include multiple choice and short-answer questions for the purposes of standardization and to bolster the psychometric properties of reliability and validity and, in turn, to provide equity parameters.

The culminating assessments would be administered near the end of the course, and be given at the school by professional school staff under standardized conditions. The goal would be to provide a basis for understanding how students at any site have done in relation to high national performance standards for *Pacesetter*. These tests would be scored by groups of teachers trained in scoring according to special guidelines.

The scoring process, in addition to serving as the vehicle to measure student learning, will also serve as a source of professional development for the teachers involved. Participating in a shared process of review, the scoring experience, as in Advanced

Placement, provides an opportunity to upgrade professional knowledge and skills.

This professional development experience will be supplemented by other opportunities for teacher training. In delivering *Pacesetter* to students, teachers will face multiple and strenuous challenges: new course content, new methods of delivery, students who are unfamiliar with this level of intellectual challenge, and new modes of assessment. Therefore, teacher development will need to be extensive, on-going and rich.

At this time, we are planning the following teacher development activities during the first two years of *Pacesetter*.

- Summer institutes to expand the tools teachers use to help all students master challenging content, such as case studies, extended concept applications and embedded performance assessments.
- On-going teacher development through workshops to review basic principles, share common experiences, and anticipate challenges.
- Technical assistance to teachers through a telephone "hotline" or possibly a mobile facilitator team to provide support and guidance to teachers in each pilot site.

Pacesetter, with its strong emphasis on formative and summative performance assessment, places the issue of educational equity and assessment in relief when viewed against the more familiar background of traditional norm-referenced tests. In *Pacesetter* a variety of assessment formats will be used. For example, multiple choice, free-response, and alternative forms of assessment will be administered and scored in a standardized manner to permit valid inferences across students and classrooms. Teachers will be working with teams of researchers from the *Educational Testing Service* to design assessment instruments that meet high standards of psychometric quality.

Again, like our strategy in *EQUITY 2000* , student achievement on these new measures will be carefully studied from the perspective of test bias and equity. Coupled with what we learn from the assessments in *EQUITY 2000*, our analyses of student

performance in *Pacesetter* courses, beginning with mathematics, will shed additional light on the role of instructional opportunities for fostering educational equity.

These two projects, *EQUITY 2000* and *Pacesetter*, represent a "push-pull" strategy. In *EQUITY 2000*, we are helping schools to "push" students particularly minority and poor students into more demanding preparation for high school, and college. *Pacesetter*, we are confident, will provide the magnetic pull toward a goal of high achievement standards for all students. We believe these two efforts, in concert, can be a major start in helping schools implement higher standards where no student is left behind. More important, these two programs--both of which feature innovative educational assessments that link instruction and assessment--are mutually supportive efforts which will lead ultimately to the goal of higher standards of achievement for *all* students.

Conclusion

The equity imperative demands an educational system in which the vast majority of students are exposed to the kind of quality education that only the top third have been exposed to up to now. And this becomes a major challenge in our public schools where the conviction runs deep that ability distributes itself along the familiar bell-shaped curve and that tracking is reasonable. Real school reform is precisely the eradicating of this belief, and installing, in its stead, not only the conviction that *all* students *can* learn, but mechanisms to realize that conviction. If future generations are to judge us favorably on issues of equity and excellence in schooling, then we must confront the wider contextual issues that will shape educational outcomes of all youth. What we are confronting is a very difficult dialectic: the desire for education based on high standards , and the equally strong need for equality of education for all. An overly strident pull toward equity can result in a tragic tumbling of all too distressingly low levels of achievement.. On the other hand, we know only too well that exclusive emphasis on high standards can result in "savagely" unequal opportunities for educational success through tracking and other discriminatory

practices.

No single innovation, whether it be *Pacesetter*, *EQUITY 2000* or whatever else comes along, whether in testing or in standard-setting, will spawn the reforms equity demands. Undoubtedly, a desired future in educational assessment will draw heavily on new theories of cognition and new technologies, reflect high standards, and foster equity. Only in this way can we be certain that all students in America really do have the same educational opportunities

Finally, equity and excellence demand a renewed collaboration among educational institutions across all levels. The key ingredients of innovation and vision are needed as well. Achieving the dual goals of high standards and equity in educational assessment will require nothing less of us.

References

Anderson, L.W. (1985). Opportunity to learn. In T. Husen & T.N. Postelthwaite (Eds.). *The international encyclopedia of education* (Vol.6, pp.3683-3686). Oxford: Pergamon Press.

Berk, R.A. (Ed.) (1982). *Handbook of Methods for Detecting Test Bias*. Baltimore, MD: Johns Hopkins University Press.

Cole, N. (1981). Bias in testing. *American Psychologist*, 36(10), 1067-1077.

Frechtling, J.A. (1991). Performance assessment: Moonstruck or the real thing? *Educational Measurement: Issues and Practice*, 10(4), 23-25.

Herman, J.L. (1991). Research in cognition and learning: Implications for achievement testing practice In M.C. Wittrock & E.L. Baker (Eds.). *Testing and Cognition*. NJ: Prentice Hall.

Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Holland, P., & Wainer, H. (Eds.) (1992). *Differential Item Functioning: Theory and Practice*. NJ: Erlbaum.

Johnson, S. (in press). *The conditional nature of equity in assessment.* Paper presented at the Ford Foundation symposium on Equity and Educational Testing and Assessment. Washington, DC.

Koretz, D. (1993). *A preliminary evaluation of the Vermont portfolio assessment plan*. RAND Evaluation Report.

Kuhn, T. S. (1970). *The structure of scientific revolutions,* 2nd Ed. Chicago, IL: The University of Chicago Press.

Linn, R.L., & Harnisch, D.L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement, 18*, 109-118.

Lehman, J.D. (1986). *Opportunity to learn and differential item functioning.* Unpublished doctoral dissertation, University of California, Los Angeles.

Maeroff, G.I. (1991). Assessing alternative assessment. *Phi Delta Kappan*, December, 272-281.

Muthen, B., Chih-Fen, K., & Burstein, L. (1991). Instructionally sensitive psychometrics: Application of a new IRT-based detection technique to mathematics achievement items. *Journal of Educational Measurement*, *28*(1), 1-22.

Payzant, T.W., & Wolf, D.P. (1993). Piloting *Pacesetter*: Helping at-risk studnets meet high standards. *Educational Leadership*, 50(5), 42-45.

Sizer, T. R. (1985). *Changing schools and testing: An uneasy proposal.* The redesign of testing for the 21st century: An ETS Invitational Conference. Princeton, NJ: Educational Testing Service.

Skaggs, G., & Lossitz, R. (1992). The consistency of detecting item bias across different test administrations: Implications of another failure. *Journal of Educational Measurement, 29(*3), 227-242.

Stewart, D. (in press). The evolution of college entrance examinations. In D.P. Wolf (Ed.), *The Ninety-Third Yearbook of the National Society for the Study of Education.*

Stiggins, R.J. (1991). Facing the challenges of a new era of educational asessment. *Applied Measurement in Education*. 4(4), 263-273.

Treisman, P. U. (1985). *A study of the mathematics performance of Black students at the University of California, Berkeley*. Berkeley, CA: University of California.

Worthen, B.R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, February, 444-454.

Wittrock, M.C. (1991). Testing and recent research in cognition. In M.C. Wittrock & E.L. Baker (Eds.). *Testing and Cognition*. NJ: Prentice Hall.

Wolf, D. (1990, June). In panel entitled: *What kind of instruction should measurement be driving?* Discussion conducted at ECS/CDE Assessment Conference, Boulder, CO.