

CHAPTER XIII

The Evolution of College Entrance Examinations

DONALD M. STEWART AND MICHAEL JOHANEK

Over the last 150 years, one of the hallmarks of American education has been the testing of increasingly large groups of people through processes of growing sophistication made possible by continuing advances in the technology of information processing. Much of this testing has been largely external to the instructional process, driven by the interests of policymakers and governments, especially vis-à-vis grades K-12, and has served various ends. A report of the Office of Technology Assessment to Congress, exploring the general history of educational testing in the United States, noted that:

- Since their first administration 150 years ago, tests have been used to assess student learning, hold schools accountable, and allocate educational opportunities to students.

- Continuous advances in design, innovation, and scoring technologies have helped make group-administered testing of large numbers of students more efficient, more reliable, and less expensive.

- Standardized tests, including college admissions tests, were perceived as instruments of school reform, and as a prod for student learning.

- Although generally viewed as instruments of fairness and scientific rigor, some educators believe that admissions tests may have exceeded the limits of their design, and more important, no longer reflect either the best thinking about how the mind works or the evolving normative goals for the inclusion of all students in the educational process.¹

One of the dynamic changes currently occurring is that educators are trying to reclaim and reshape educational assessment, including admissions testing, to serve purposes of teaching and learning as well as monitoring and accountability. One outcome of the change in the way tests are viewed is that “performance-based” or so-called “authentic” assessment—examined in various ways by other authors in this

Donald M. Stewart is President of the College Board. Michael Johaneck is a Research Associate and Special Advisor to the President at the College Board.

volume—is being proposed as the basis for a new wave of admissions testing. Performance-based tests may include writing samples, extended constructed response items, portfolios of student work, exhibitions, simulations, and interviews as means of supplementing traditional testing methods. The College Board's efforts along these lines, evident in its evolving Pacesetter initiative described below, illustrate the promise that performance assessment offers in the pursuit of educational reform. Future technological breakthroughs offer further exciting opportunities, and their possible directions for admissions testing are outlined in this chapter as well.

But innovations in assessment formats, as even their most fervent advocates will testify, do not in themselves answer the more fundamental questions of the ends of education, of the normative values we wish these assessments to serve. Whatever system of assessment we use, it will reflect the fact that education is not random. It is animated by desired processes and it proceeds toward defined ends and values. One such fundamental value of the College Board, as evident in its EQUITY 2000 initiative described in this chapter, was expressed by a distinguished director (and philosopher) of college admissions, B. Alden Thresher some twenty-five years ago: "To a much greater degree than anyone realized, talent is not something stumbled upon and found here and there; it is an artifact . . . [and] it can be produced."² An underlying supposition of this chapter is that talent can be produced, *all* children *can* learn, irrespective of their backgrounds, socioeconomic situations, race, or gender. As a result, the primary goal of education, which testing must support, is educational equity for all students, the right of every student to receive the same high quality education that previously has been reserved for only the top quarter of the population.

In order for that to happen, we must confront the wider set of forces that shape both the design and the uses of assessment instruments. It is here that history can offer us some limited though critical counsel, to which we will turn later in this chapter. We may be well advised to hear the optimism of a past reformer of testing, Professor Edward L. Thorndike of Teachers College, Columbia University, writing in 1923 in the 21st Yearbook of the National Society for the Study of Education. After dismissing concerns that the new science of educational measurement would simply encourage students to work for grades and hinder good teaching, Thorndike addressed a final concern:

It will be said that . . . the finer consequences for the spirit of man will be lost in proportion as we try to measure them, and that the university will become a

scholarship factory, turning out lawyers and doctors guaranteed to give satisfaction, but devoid of culture. . . . [T]he fear is groundless, based on a radically false psychology. . . . Of science and measurement in education as elsewhere, we may safely accept the direct and practical benefits with no risk to idealism.³

We are not sure we share such optimism today. The broad experience of the College Board with the many facets of the transition from high school to postsecondary schooling provides it with both the enthusiasm to pursue the promise of performance assessment and the larger realization that any component of reform must directly address the wider context that will likely shape the values it ends up serving.

In this chapter we first examine the current state of the art of testing in light of the call for educational reform and then examine new initiatives being taken at the Board through an integration of assessment and teaching. Second, we peek into the future where technology provides startling possibilities for assessment to enhance individual learning and personal productivity. Finally, the essay addresses the counsel of the Board's own history and lays out several of the key themes that have dominated the last century of admissions testing in the United States.

The Current State of the Art

For the past five years, some political thinkers have noted that what civil rights began at mid-century, economic forces will attempt to complete as we enter the 21st century, namely, the full educational preparation, and economic "mainstreaming," of members of all ethnic and racial groups. In addition to being the right thing to do morally, in their view, it is also unquestionably the necessary thing to do if we are to have a society that can sustain a high quality of life for all and remain competitive in the world economy.

Moreover, the agendas of the Clinton administration, of several governors, and of some policymakers have called for voluntary, national educational standards with the integration of policies and programs around those standards. Within this context, the use of assessment will take on multiple purposes: to help inform instruction, to provide data on the success of schools and systems, and to refine the delivery of education in order to help all students meet those goals. As part of this effort, the battle will be joined between (a) those teachers, administrators, and others who feel constrained by tests that do not help them reach educational goals directly and (b) other educators and

leading educational measurement experts who emphasize that quality standardized tests are useful tools in gauging the strength, weakness, and progress of American students.

The College Board has been involved in this debate for well over a decade. A long-term decline in the average SAT scores, which began in the 1960s and was recognized and studied in the 1970s by the Wirtz Commission,⁴ was also cited by the U.S. Secretary of Education, Terrell Bell, in his 1983 report as one of several indicators that we had become (as the report was entitled) *A Nation At Risk*.⁵ (Unfortunately, the neglect of educational and social needs generally can be seen in the gyrations of the economy as well as in disastrous social upheavals from Los Angeles to Crown Heights.) From one point of view, the comparability of the SAT over time—the possibility of comparing levels of learned, precollegiate verbal and mathematical abilities among a key group of students over different years—has provided moderation of the sad practice of grade inflation in high school and has given an indication of the decline in the level of preparation of students for college. It is significant that in the mushrooming of “developmental courses,” much of American higher education today resembles the community college sector of two decades ago.

We must, however, go beyond mere measurement of our changing levels of ability, and devise programs to improve them. Also in the early 1980s, as part of its groundbreaking work in determining, through national consensus among educators, what students need to know and be able to do in order to go to college, the College Board canvassed business leaders about their requirements for entry-level workers. They, in turn, told the College Board that the basic precollegiate skills are also those needed for entry directly into the workforce. (This work culminated in the *Academic Preparation for College Series*, a set of six booklets describing the knowledge and skills needed by all college entrants in the basic academic subjects.) As jobs have become more sophisticated, the ability to think independently, work collaboratively, and communicate effectively, as well as the achievement of collegiate levels in mathematics and verbal abilities, is becoming widely recognized as requisite for entering the twenty-first century workforce as well as postsecondary education.

The consequence of this is that an educational system based on tracking by presumed ability levels in the early grades, followed by the filtering out of the top third of the students for postsecondary study, must give way to a new philosophy. This new philosophy, this new normative view, of what we want education to do, asserts that all students

have ability, all students have the right to a quality education, assessment must inform and support instruction, and college entrance examinations must both reflect and reinforce this new philosophy. Therefore, voluntary national education standards must be defined and appropriate assessment instruments identified or developed to measure the progress of students and schools in meeting these standards.

New Initiatives by the College Board

Had the College Board developed new forms of assessment in the 1980s to match the work done in the *Academic Preparation for College Series*, the nation would have had a good jump on the kind of assessment now being called for. Responding to social and political change, as well as to the specific mandate of its trustees, the Board in the 1990s is moving to answer these needs. Current cognitive theory and curriculum reform are built on the view that learning is constructive and interpretive and that learning increases when knowledge is reconfigured and facts and skills are interpreted in relevant contexts. As a result, assessment is changing to reflect how students think, perform, and learn from instruction.

These developments in cognitive theory and curriculum reform illustrate that our understanding of how people think is fundamental to education and to assessment. Responsive to these changes, the new College Board initiatives comprise the revising of the SAT to reflect curricular trends and to include "performance-based" elements; expanding the use of Advanced Placement courses and assessment to enrich the high school experience among poor and minority students; and the creation of three major programs of precollegiate preparation and assessment—the "new" SAT, EQUITY 2000, and Pacesetter. These programs, starting as early as the middle schools, are designed to end the pernicious practice of tracking and to promote significant and systemic school reform, in order to insure that all students are ready for college whether they choose to continue their studies immediately after high school or enter the workforce directly. It is impossible to understand the Board's larger philosophy about the transition from high school to college, and thus the role and evolution of entrance examinations, without taking these new programs into consideration.

THE REVISED SAT

The "new" SAT, introduced in 1993/94, reflects technical and philosophical changes as well as what students experience in today's

classrooms. SAT-I: Reasoning Tests requires the familiar three hours of testing time, with some significant changes. In the verbal section, antonyms no longer appear. In their place are more reading passages that reflect both what colleges expect of students and current instructional theory, focusing on the student's ability to read critically. Thus, approximately half of the questions are based on longer, more engaging passages to be read and include a pair of passages on the same or related topics, one of which opposes, supports, or in some way complements the point of view of the other. Even with vocabulary questions, the emphasis is on testing students' verbal reasoning skills and knowledge in context.

Similarly, on the mathematics sections of the SAT-I Reasoning Test the focus is on problem-solving skills important for success in college, and there is an increased emphasis on a student's ability to apply concepts and interpret data. In addition, for the first time, there are questions that require students to supply their own answers instead of selecting one from multiple-choice alternatives, and students are advised to bring a calculator.

The SAT II: Subject Tests, still known as the Achievement Tests, reflect important new directions and give students an opportunity to show their academic strengths in a wide variety of subject areas including writing, literature, foreign languages, history, mathematics, and sciences. New tests in Japanese and Chinese include a listening component. A Korean Achievement Test is also being created. Tests with listening components, including new offerings in French, Spanish, and German, are administered in secondary schools, not in test centers. Although not officially launched until 1994, some changes were introduced earlier and the introduction of new elements in the SAT II test will continue throughout the decade.

The new Writing Test includes a direct writing sample and questions that require recognition of the conventions of standard written English, as well as effective and logical expression. Among the advantages of the test are:

- an essay providing a direct measure of writing ability while not assuming any specific subject-matter knowledge;
- revision-in-context passages that present a context larger than a discrete sentence and therefore permit questions on logic, coherence, and organization which are similar to common in-class exercises in which students revise their own essays; and
- usage questions requiring students to recognize errors, and sentence-correction questions requiring recognition of errors and selection of the correct rephrasings.

It should be noted that, despite the inclusion of the word "aptitude" in its original name, experience over the last seven decades has shown that scores on the SAT reflect learned abilities rather than some inborn, immutable level of intelligence. Accordingly, the SAT is now called the Scholastic Assessment Test. This confirms Thresher's opinion that "talent can be produced" and that student abilities are a function of work done by the individual. Interestingly, as several observers have noted, schooling also appears to raise general developed cognitive ability as measured by tests of what was thought to be an inherent quality (intelligence) as well as by so-called aptitude tests. Thanks in part to the rise in educational attainment throughout the United States, the average IQ has been rising for much of the twentieth century, a phenomenon found in other nations across the globe.

Similarly, there is a strong relationship between academic preparation and achievement on the one hand and mean SAT scores on the other. Over the past decade, the College Board has reported a consistent pattern: the more years of academic study and the higher the grades and rank-in-class achieved, the higher the SAT score. In 1994, for example, students who reported taking physics in high school had average SAT verbal scores of 463 and average mathematics scores of 538, considerably *above* the national average for each. (Students who took calculus in high school had the highest average in mathematics, 598, *and* the highest SAT verbal average, 501.) By contrast, students who had taken mathematics courses other than those in the traditional college preparatory sequence had SAT scores 23 points *lower* than the national average on the verbal section and 31 points lower on the mathematics section. This clearly contradicts the assertion of some critics that the SAT is unrelated to course work and thus is an unfair or inappropriate measure of student abilities. On the contrary, the clear statistical evidence is that the more and better courses students take, the better, on average, they are likely to do on the SAT as well. (Since, by and large, certain minority students are still being tracked out of academic courses in the early grades, it is not surprising that as a group, many minorities do not do as well on the SAT. The issue is one of "savage" lack of equity in schooling, not discrimination by the test against one minority group or another.)

Admissions testing is just one step in a continuum of awareness, learning, and preparation that begins well before entrance to college. If all students are to reach new high standards at graduation from high school, efforts toward higher standards of preparation in earlier years

(middle school and early high school) will also be required. The College Board has been concerned about the extent to which mathematics has become a filter of students in the nation's high schools. Those who take algebra and geometry have the option not only of higher mathematics, but of 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. Thus, we started a project called EQUITY 2000.

EQUITY 2000

EQUITY 2000 is a national school reform initiative designed to achieve the following ambitious goal: By the end of the twentieth century, minority and disadvantaged students will enroll in and complete college at the same rate as majority students.

Three elements underlie this initiative. First, the program ensures that *every student* in the participating school districts completes algebra and geometry, the "gatekeeper" courses that are prerequisites to preparation for college. Experience and research indicate that students who enroll in college preparatory mathematics courses are likely to enroll in college preparatory courses in all subjects. According to a study by Pelavin and Kane, high school students who took one year or more of algebra were two to three and a half times as likely to attend college as students who did not take algebra.⁶ Students taking geometry and algebra are between three and five times as likely to attend college as those who have not taken either.

Second, EQUITY 2000 involves teachers, parents, counselors, principals, and others in the educational community to create an ethic of educational excellence within entire school districts.

Third, the program uses a variety of activities to enrich students' academic experiences, particularly in mathematics, and to build aspirations toward a college degree.

The program represents the College Board's commitment to districtwide (K-12) systemic change in school districts, and is being implemented in fourteen districts through six EQUITY 2000 sites around the country. The model includes the elimination of tracking—a policy that leads too many students to an academic dead end—and also includes rethinking the way to handle heterogeneous mathematics classes. With students having widely varying backgrounds and experiences in mathematics, heterogeneous classes require that teachers approach mathematics instruction in a manner that enables every student to achieve his or her fullest potential.

EQUITY 2000, for the first time, executes a full complement of precollege initiatives within entire school districts so as to reach every student. Four major components comprise the EQUITY program: in-service training for teachers and administrators; academic enrichment for students; parental and family involvement; and community involvement and support.

In-service training for teachers and administrators. Teachers, counselors, and principals are all crucial to changing educators' expectations of students' success. At summer and academic year institutes, mathematics teachers strengthen their knowledge of mathematics and enhance their ability to teach effectively in heterogeneous classes. Guidance counselors focus on strategies for building students' aspirations to pursue a college degree, including advising students to enroll in college preparatory classes, involving parents in students' academic development, and disseminating information about college options and costs to *all* students. Principals focus on creating a wholly supportive academic environment in their schools for all students, including the creation of teacher-counselor teams that meet the needs of individual students. Far exceeding our expectations, more than 2400 teachers, counselors, and principals participated in the 1994 summer institutes and workshops.

Academic enrichment for students. Because academic tracking has been so pervasive and has denied the opportunity of algebra and geometry to so many students, schools have an obligation to offer them academic enrichment activities to ensure that they participate in the courses on a "level playing field." These activities, aimed at raising both skills and expectations, include:

- Summer Scholars Programs. Taught by teams consisting of a college or university professor and a master teacher from the local school system, these programs are designed to support regular academic course work every day for five weeks during the summer;
- Saturday Academies. Six Saturday sessions are similarly team taught and give attention to skill-building across many disciplines; parents accompany their sons and daughters to class and work with them on specific academic exercises;
- Academic Enrichment Laboratories. Capitalizing on the impact those nearest to the students' age group can have, these laboratories use college and university students as role models for middle and high school students.

Parental and family involvement. Family involvement is critical. Two specific activities in EQUITY 2000 are (a) family mathematics workshops and other activities in which parents participate jointly with their sons and daughters in academic work and college planning; and (b) Career/Resource Centers at each school, through which information on college and career options is disseminated.

Community involvement and support. Each site has established formal partnerships with institutions of higher education and with community organizations to support and participate in EQUITY 2000 program activities, such as the team teaching at all Saturday Academies held on local college/university campuses.

The summer institutes and follow-up workshops for mathematics teachers do not occur in isolation; a serious commitment to both equity and excellence would not allow it. Rather, they are key components of a comprehensive educational reform initiative that includes a variety of interventions leading to districtwide (K-12) systemic reform. The program is being thoroughly evaluated by an independent research group, and the College Board intends to disseminate the model broadly to school districts across the country using its vast network of resources.

PACESETTER COURSES

From the creation of the SAT in 1926, which made it possible to identify a significantly larger pool of qualified candidates for college, to the subsequent use of the SAT to attract a considerably larger number of minority students to higher education, to EQUITY 2000, whose goal is to put all students on a full academic track, the College Board has contributed to an increased democratization of education in America. In this same spirit, another new effort of the College Board—the Pacesetter courses—is directed at helping schools implement higher standards by providing new course syllabi and assessments for high school courses in key subject areas, through which teachers can learn to teach to higher standards as well. These courses are being designed so that schools that wish to raise their standards will have course materials, assistance with preparation of teachers, and assessments to help them achieve that goal. In most cases, Pacesetter courses will embody the top level of the high school curriculum. By contrast, Advanced Placement courses, for which Pacesetter will prepare students, are essentially college-level courses given in high school to those students ready to take them.

Based on high standards, Pacesetter initiatives take as their starting point the development of detailed substantive course frameworks each of which specifies the structure and content of what should be taught and learned in a key course of study. Pacesetter will offer professional development activities so that teachers will be able effectively to instruct all students, as well as activities in the classroom that allow students to translate concepts into hands-on situations, and teachers to evaluate students' skills and understanding, thereby informing instruction. There will also be end-of-course assessments to measure student attainment of course objectives.

Pacesetter courses in mathematics, English, and Spanish are currently in field tests; subsequently we intend to complete development of world history and science courses. 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 Studies; the National Science Teachers Association, and the American Council on the Teaching of Foreign Languages.

As we move forward with the process of educational reform, we must keep in mind the distinction between assessment for instructional purposes as opposed to accountability purposes. 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 facilitator and mentor.

By contrast with the in-course 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, and validate student accomplishments. In addition, they will 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 each of the Pacesetter Culminating Assessments as being 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 subtasks 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 provide a brief reflection on the integrating process.

Thus the Culminating Assessment would call for (1) 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 needed for the purposes of standardization and comparability.

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. With Pacesetter we are providing a concrete pull toward a goal of high standards of achievement for all students. We believe these two efforts together can be a major start in helping schools implement higher standards in programs where no student is left behind.

As exciting as these innovations are, we must remain circumspect. After studying initiatives like the alternative assessment system being pioneered in Vermont, a growing number of researchers are sounding cautious about the use and the cost of performance-based assessment. The nature of the tasks, inequalities surrounding the opportunity to learn the requisite skills, and the unreliability of scoring procedures all appear to present potential for disadvantaging minority students and offer serious challenges. Recent Rand Corporation reports found that the reliabilities of the portfolio and performance-based assessment among Vermont fourth and eighth graders in both mathematics and writing were quite low, although positive with regard to the professional development of teachers.⁷ Researchers in other situations have

found problems with regard to mathematics in insuring reliable scores caused, in part, by the prohibitively large number of tasks a student needs to perform in order to get reliable estimates of his or her performance. It is, of course, wonderful to believe that we have a new system that will solve all our old chronic problems. But it would be foolish to embrace what, in fact, is not yet fully substantiated. As researchers Dunbar, Koretz, and Hoover put it: "Quality control in terms of both evidence and consequences of (performance assessments) is not a question of faith, but an empirical matter when measurement is intended to inform public policy."⁸ Decisions, therefore, about the kinds of assessment to use in the future might be based not only on an accurate analysis of current forms of assessment, but also on a complete and well-documented evaluation of the proposed options.

The initiatives described above have been inspired by new cognitive theories and are being developed using current technologies in testing. Yet, perhaps the greatest potential for creating assessments that truly empower students and support the process of learning will come from future advances in technology and psychometrics.

*Technological Advances in Assessment:
A Challenge to the Future of Assessment*

As we look to the year 2000, we find ourselves on the verge of a revolution in educational testing and assessment. A number of forces, some societal and others technological, are the engines of change, compelling us to look carefully at how and why we test in our schools and colleges.

Currently in its embryonic state, new test theory promises to influence and shape the practice of educational assessment and make possible a host of new educational measurements. In addition to continuing to capitalize on the advantages of multiple-choice-type items, future assessments will include, among other things, extended responses constructed by students, portfolios, and simulations of scientific experiments and modeling. As is true of well-constructed multiple-choice questions, these new kinds of assessment will require higher-order problem solving and critical reasoning. Therefore, college entrance testing programs, such as the SAT, undoubtedly will continue to undergo dramatic redesign in the twenty-first century.

The revolution in the cognitive sciences beginning in the 1950s and 1960s, for example, has brought us a new view of learning. No

longer do we view students as passive learners and teachers as “talking heads.” Today’s notions include the learner as an information processor, the teacher as a facilitator of learning and not merely an expert dispenser of facts and prescriptions. Technologies emerging from the field of artificial intelligence are now beginning to adopt many of the advances in cognitive theory. Not only are we witnessing the development of computer systems that “understand” spoken language, programs that mimic “experts” and “novices” in particular fields or specialty areas but also systems that “learn” from experience and data. Many believe that in the not too distant future testing will permit us to assess not only traditional forms of knowledge, such as the acquisition of facts and analytic abilities, but the very nature of the learner’s “mental models,” i.e., how he or she views a complex system of concepts and rules. More important, these new assessments will help us design and deliver instruction that serves to correct the inaccuracies and incongruities in those mental models. Although we must be vigilant that new technologies do not exacerbate the differences in opportunities that exist currently among different groups in our society, smarter testing technologies, assessments that adapt to the examinee and provide enriched diagnostic information, will offer the opportunity to tap new dimensions of developed ability. This new assessment paradigm, one that includes intelligent testing technologies, will foster a fundamental change in the nature and uses of educational tests and assessments.

New forms of assessment, such as adaptive testing, portfolios, and other types of performance assessment, will continue to be developed to measure both individual achievement and the effectiveness of local educational reforms. Currently, many states and school districts are placing an increased emphasis on performance assessment as a means of supplementing their traditional testing programs. As students, teachers, parents, and educators gain more familiarity with these methods, and as the test theory and psychometric methods needed to ensure quality and fairness are advanced and developed, this trend in educational assessment will continue to develop if issues of cost and feasibility can be resolved.

The most likely development is the powerful combination of advances in computer technology with respect both to hardware and software, and progress in test theory, that will lead to a widespread increase in the availability of different kinds of tests. Just as increasing access to higher education since World War II accomplished a democratization in education, these new tests will accomplish a democratization of assessment. Changes of this sort will place the information

derived from educational assessments more directly into the hands of students, teachers, administrators, and policymakers, and enhance the instructional relevance and desirability of individually administered tests. Current projects to increase the uses of computer technology, including the development and construction of powerful national, and ultimately global, computer networks, as well as an abundance of more powerful and affordable personal computers suggest that computer-based testing products and related services will grow exponentially in the next decade and beyond.

With its long history of sensitivity to the needs of test takers and test users, the College Board is moving actively in the direction of increasing the role of computer-based technology in its testing programs. These new directions and initiatives are captured by the Board's project tentatively called *Transition 2000*. Our blueprint for future assessment will move beyond paper and pencil tests and operate on a common or linked delivery system. Much like the electronic databases commonly used today, students and their families will be able to create, maintain, and access computer-based academic portfolios which will include transcripts, applications, and assessment information. Ease of access and encrypted file transfer will make transmission of all or parts of a student's academic record to colleges, universities, and other institutions of higher education easier for both students and school administrators. The burdensome paperwork, long delays, and the attendant uncertainty of applying to college will be a thing of the past.

The range of potential enhancements in the assessment include:

- the use of adaptive testing to reduce students' frustration with too easy or too difficult tests;
- the widespread use of performance-based assessments, like portfolios, simulations, and extended response formats;
- closer links between measures of reasoning ability and academic achievement;
- the introduction of new measures of affect and of cognitive abilities (motivation, practical intelligence, and learning abilities);
- an increase in the diagnostic information available to students, parents, and teachers to aid in guidance and instruction.

The long-range success of the *Transition 2000* project depends on finding effective ways to deliver these new services to all students. Thus, we will be exploring the role that can be played by high schools, community-based computer centers, colleges, and private businesses.

There is no doubt that in the future microcomputing technologies will permeate the school environment. As already envisioned and piloted by producers of hardware and software, master classrooms complete with multimedia tech-desks and liquid crystal displays will become the teaching tools of a new generation of teachers and the learning environments of our grandchildren. Students will take notes using electronic organizers, personal computer assistants, and other forms of microchip "techno-tools," many of which have yet to be designed.

Intelligent assessments like the SAT will include multimedia presentations, brilliant color palettes, lively animation, and 3-D graphics that will bring simulations and other forms of dynamic, interactive testing to life and help make assessment more authentic. Test formats that include choosing from among an array of options using pull-down menus and windows, conceptually similar to multiple-choice tests of today, will be used to gauge academic abilities accurately and fairly. After taking the next generation of "smart" tests and using the electronically linked guidance software, students and adults will have a clear idea about levels of achievement and what they need to do to gain admission to college and to succeed there. Testing in a high-tech educational setting will be radically different from what we have in American schools now. In the future, the SAT may well be available on demand at the local high school's computer laboratory or at a learning center or public library.

In sum, the "black box" of testing technology, formerly controlled by remote "experts," will be opened for full participation and extended use. As a result, democratization of assessment can be expected to expand well beyond its current scope. Drop-in, on-demand testing will be the norm everywhere. With test scores in hand and detailed information about their proficiencies in specific skills and academic subjects, students will leave the testing situation with the intellectual capital of self-knowledge. They will be armed with the information to make meaningful choices in the near term, choices that will help secure access to higher education in the future.

Tests and other educational assessments—heretofore only seen and appreciated by a handful of scientists and educators—will be seen as fair, user friendly educational tools, an array of powerful instruments for self-transformation and educational change.

Lessons from a Century of Testing

The latter part of the twentieth century is witnessing an explosive growth in testing in an effort both to stimulate educational reform and

to capture its successes. With the current emphasis on national education standards and the advent of widespread computer-based testing, the evolution of educational assessment will undoubtedly continue well into the next century.

For many—students, teachers, and parents, as well as testmakers and psychometricians—this ongoing evolution is a welcome development based as it is on advances in cognitive theory and psychometrics. In the future, we can expect to know more about how humans learn and what schools can do to foster learning. To keep pace, the science of educational measurement will push test design and construction into new and, we hope, more productive directions. So-called “smart” computerized testing technologies may help further the ongoing democratization of the educational process.

Over its 93-year history, the College Board has worked with countless changes in testing. From this experience, it may be useful to ask how assessment changes have fit into the wider web of forces in which schools are embedded. A glance backward may help us to move forward, and to do so with a refined view of the complex forces shaping educational outcomes, including those forces shaping the ends served by assessments.

The College Entrance Examination Board was brought into being on the campus of Columbia University in 1900 under the tutelage of Harvard president Charles Eliot and with the leadership of Columbia’s president-to-be, Nicholas Murray Butler. The goal was to resolve what was then called “educational anarchy” caused by the fact that each postsecondary institution had its own admission procedures at a time when the number of public high school graduates had more than quadrupled over the previous two decades. A common examination, cogently advocated by President Eliot as early as 1890, was the solution agreed upon by the small group of forward-looking representatives of schools and colleges who founded the Board. Originally, this examination was a series of what we would today call performance-based tests. The “College Boards,” as they soon came to be called, were subject-matter essays which students wrote in “blue books.” The problem was that this approach assumed a specific curriculum which only a few elite preparatory schools followed. After World War I, during which a multiple-choice, general abilities model of testing Army recruits was developed, the College Board devised a new admissions test, the SAT, using a new multiple-choice format. Far more curriculum-neutral than the original College Boards, it allowed colleges to identify qualified students from high schools all across the nation no

matter what the specific nature of the curriculum might be. In so doing, it greatly served the democratization of higher education based on academic merit by making it possible to identify qualified students from secondary institutions that were remote from, and unknown to, the top tier of colleges and universities.

The historical context of testing over the last century provides important insights into factors beyond the classroom doors that have influenced—and may yet influence—the evolution of testing and the aims it serves. At least four sets of factors have significantly shaped the design and implementation of testing:

- national security interests and the demands of wartime;
- changes in psychological theory and in the testing industry's technology;
- changes in organizational theory and practice in business;
- developments that have affected the “politics of knowledge.”

External forces continue to influence how we think about testing, and must be addressed if testing innovations are to serve the goals of equity and excellence that we desire.

National security interests. Curious as it may seem at first blush, throughout this century national security interests and the demands of wartime have spurred waves of innovations in assessment. With the goal of sorting top talent for the common defense, large-scale, systematic testing of individual aptitudes and achievement levels first occurred in the United States under the pressure of World War I, with the famous Alpha-Beta tests administered to over a million recruits. This effort gave a tremendous boost to measurement psychologists, and while the test results did not inspire much enthusiasm within the Army, a corps of several hundred psychologists saw the possibilities for wider use. After the war, they promoted more systematic and more “scientific” tests, finding a receptive audience among educators. By 1920, some 200 colleges and universities were administering Army Alpha tests or similar instruments, especially for purposes of college admissions. To develop what became the SAT the College Board in 1924 chose a group of psychologists, including many like Robert Yerkes and Carl Brigham, who had played key roles in developing the Army tests. Two years later, over 8,000 students took the first SAT. While the enthusiasm for mental testing often overran the technical capacities of those early tests, and while much skepticism arose during

the testing boom of the 1920s, the use of standardized testing took significant root in schools across the United States.

World War II provided another key boost in testing. The Air Force developed a whole series of special batteries for pilots, radio operators, range finders, and other military specialists. For more generalized ability testing, the Office of Strategic Services (OSS), the precursor of the CIA, developed a performance testing program for selecting spies that later inspired assessment methods for business executives.

World War II also forced retrenchment in school testing; the demands of the war were cited as the chief reason for dropping the essay-type exams the Board had offered for forty-one years. Owing to the need to provide more efficiently for the personnel needs of the armed forces, colleges sought to start first-year courses during the summer, requiring a faster turnaround time for reporting examination results than the essay format would allow. Wartime economies also encouraged cuts in the costs of the examinations. The multiple-choice SAT, along with the recently developed multiple-choice Achievement Tests, then became the norm for subsequent decades.

The Cold War, especially after the Korean War, provided its own impetus for better assessment techniques. In an era keen on human resource management, Henry Chauncey, president of Educational Testing Service, his friend James Conant, and others left little doubt that improved school testing and guidance formed the core of an efficient process for identifying the more accomplished students that was critical to the national interest. The number of students taking the SAT grew dramatically during this period, rising tenfold between 1951 and 1961. The Educational Testing Service (ETS), it should be noted, was founded in 1947 by the College Board, the Carnegie Foundation, and the American Council on Education in order to handle the increasingly complex and psychometrically technical aspects of preparing the SAT and the Achievement Tests and to oversee their actual administration. Over the past three and a half decades, ETS and the College Board have worked closely together with regard to the SAT and to other testing and financial aid programs as well.

In the post-Cold War era, uncertain as it currently is in geopolitical terms, we are told that the battlegrounds will be economic, and the shock troops will be well-trained workers. For the United States to prevail in an "economic conflict," all our human resources must be fully tapped. Whereas Sputnik fueled reforms in mathematics and science to compete against the Soviet Union, international economic

competition inspired a new wave of reform, with generals marching forth from the business community, announcing that if we are to maintain economic security, we must improve our ability to develop all talents and assess them accordingly.

Advances in psychology and testing technology. More obvious forces influencing the evolution of testing are ongoing advances in psychological theory and in the testing industry's technology. Psychologists' work on intelligence and cognition theories has continuously informed changes in testing—from Galton's studies of heredity, to Binet's "mental level," to Thorndike's "connectionism" and scales, to Spearman and "factor analysis," to Cattell's fluid and crystallized intelligences, and to the more recent work in cognitive psychology by researchers such as McClelland, Resnick, Hunt, Sternberg, and Gardner. More recent advances in understanding cognitive functions have encouraged educators to develop assessments for these processes and to abandon the pedagogical assumptions of outdated learning theories. Assessment and instruction must reflect new understandings of how we process information and how we construct meaning.

Advances in testing technology inspired by the pressures of national security as well as by changes in psychological and psychometric theories have facilitated an ever larger number of test-takers and established a vast testing industry. Industry suppliers and educational psychologists frequently have cooperated in implementing testing innovations. For example, in 1928, the expense of hand scoring early tests prompted Columbia psychologist Ben Wood to urge ten corporations to develop a scoring machine. A response from Thomas Watson of IBM led to a lifelong association between Wood and Watson. An IBM tabulator later reduced the per-test scoring cost of the *Strong Vocational Interest Blank* from \$5 to 40 cents, encouraging wider usage of the test. Recent advances in multimedia computer technology promise to facilitate broadened assessment methods, with the eager support of the computer industry.

Changes in business organization. Another influence on the development of education and testing has been the change in management theory and business organization, which has often reverberated through the administration of schools. In particular, shifting notions of productivity and efficiency have been used to justify the aims and uses of testing. Sensitive to their relatively low status in communities and faced with considerable managerial challenges, school administrators eagerly borrowed both ideology and terminology from the business community.

This appears to be a recurrent theme in schools from the scientific efficiency days of Frederick Taylor to the more recent PPBS (program planning and budgeting system) enthusiasm, to "flatter organizations," or the current total quality management (TQM) drive. The current emphasis on testing the *process* of schooling—consistent with the wide interest in quality management principles—clearly reflects business management's continuing influence. The uses of testing have been affected by this historical relationship for better and worse. Valuable assessment innovations born of dreams of efficiency have often fallen prey to our desire to sort human lives based on perceptions of their probable destinies, or to subordinate schooling's democratic promise to other, less enabling, economic interests.

The politics of knowledge. Finally, the past century has seen a number of developments in testing that have been affected by what has been called "the politics of knowledge." This phrase, coined by Ellen Condliffe Lagemann, refers to a set of questions including: What knowledge is authoritative? Who determines this? How does one gain access to this group? And how do the "experts" communicate with the non-experts? Part of the appeal of recent proposals for new forms of assessment has been their claim to facilitate wider educational reform. But if so, we then must ask, How will we produce authoritative knowledge about new kinds of assessment, and how will this affect the roles of both institutions and individual practitioners? Innovations such as performance assessment imply changes in how authoritative knowledge about testing is produced and distributed, and these changes need to be discussed and debated. Moreover, the challenge of maintaining standards in more variegated testing formats forces into consideration substantial changes in the politics of knowledge with regard to assessment, which need to be addressed directly and explicitly.

A brief historical comparison from the early years of the SAT may be illustrative. When the SAT was undergoing initial refinements and debate in the early 1920s, the original College Boards, the subject-matter essay tests, were still being given for purposes of admission. At the same time, a group of thirty "progressive schools" under the aegis of the Progressive Education Association began its Eight Year Study of student success in college as it related to admission testing procedures. On one level, these three simultaneous admissions approaches—the still-fledgling SAT, the traditional "boards," and the arrangements of a set of progressive schools with various colleges—represented three different assessment systems. On another level, each approach reflected a

distinct politics of knowledge, that is, they reflected different answers to the questions of what knowledge was authoritative and who determined this knowledge. The original "boards" were largely created and graded by university professors, with a few high school teachers assisting, and were meant to assess student achievement in subject areas determined by the colleges. A group of progressive schools developed their own means of assessing student performance based on their own vision of education in which high school teachers carried out evaluations and submitted them to colleges that had agreed to participate. The SAT, developed to assess "aptitude," was both created and evaluated by teams of university-trained psychometricians. The choice of which assessment model to use involved implications far beyond the technical pluses and minuses of the specific testing techniques employed. It was based, rather, on which group mounted the best claim to "authoritative knowledge." The subsequent triumph of the SAT had to do not only with its utility and practicality, but also with how these issues related to the politics of knowledge were sorted out. Similar to the rise of the other professions in the United States around the beginning of the twentieth century, testing expertise was also seen as residing not primarily among teachers either at the secondary or postsecondary level, but rather among a central core of university-trained professionals, including psychometricians, who were acknowledged as possessing authentic knowledge of educational measurement. Thus, it was their right and responsibility to devise educational programs and instruments of assessment.

Narrower institutional politics may have also played a role. As late as the mid-1950s, College Board membership included 172 colleges and 24 associations but no high schools. Only during the late 1970s was a rough parity between schools and colleges established, a parity that became more evident in the late 1980s, when community colleges began to find a place within the Board as well. It is only natural that a university-dominated College Board would emphasize the predictive roles of testing over its potential to inform instruction.

Today, we see a shift in point of view. Current Board efforts to enhance assessment's instructional role, as in Pacesetter and the new SAT initiatives, reflect the Board's more balanced school-college membership and more collaborative organization. Not surprisingly, the Board has developed a new mission statement committing it to "educational excellence for all students . . . through the ongoing collaboration of schools, colleges, educational systems and organizations." The implications of this commitment for student preparation and assessment are far reaching.

Brief as it is, this historical reflection should help inform current debates about assessment. Any innovation in assessment must take into account the social, cultural, and institutional factors that will influence its implementation. If we value equity, we must face directly how these contextual forces are likely to affect the equitable distribution of educational outcomes. We must realize that any change in a factor so central as testing is not neutral with regard to purposes and ultimate aims. Technical adjustments to the assessment system do not answer the questions about why we are assessing what we are assessing, or who is determining the ends of testing. The premises from which a particular assessment develops may not be the purposes for which it is employed. Testing reforms in the past have not only been shaped by their designer's intentions; they have also been shaped by perceived national goals, economic interests, and institutional constituencies. Therefore, to assume that performance assessment in and of itself will spur the kinds of far-reaching changes in schooling that many claim for it is to ignore crucial lessons from our history as well as recent experience in practice.

Over the last century, we have seen the influence of a variety of exogenous factors such as national security interests and organizational changes in business on the aims and implementation of testing. Prudence urges us to consider such factors in the present and future as we evaluate the impact of proposed changes in assessment on the goals we hold for our children's education. It is in this spirit that the College Board eagerly pursues the promise of performance assessment in its Pacesetter and Transition 2000 initiative, while insisting through EQUITY 2000 on addressing the contexts that will shape the ends assessment serves.

Conclusion

The College Board's experience in addressing the admissions process—what Alden Thresher called the “great sorting”—reinforces the imperative that we must continue to confront the contextual and historical factors that have directly influenced testing. While the College Board's initial years concentrated on its subject-matter essay tests (“The College Boards”) and later on the multiple-choice based SAT, in the post-World War II years the Board's involvement has expanded to include financial aid, academic enrichment, equity, and public policy initiatives all in the service of improving the transition from high school to college. The Board's current work with detracking through

the Equity 2000 and Pacesetter projects illustrates the need to address the entire schooling system in the effort to enable all students to succeed. If we are truly serious about such goals as equitable educational outcomes, we must work together in addressing the social and educational factors, including test innovations, in the reality of today's environment.

Continuing immigration, the growing disparity in the distribution of wealth, international economic competition, and the search for a new international order echo worries expressed a century ago in a very different world. At that time, many people also called for the dramatic restructuring of schooling as the twentieth century approached. In 1894, the Committee of Ten attempted to indicate, among other goals, "the best methods of instruction" and "the best methods of testing the pupils' attainments therein."⁹

While the Committee's report received a great deal of attention at the time, its direct influence on schools appears to have declined dramatically by the early 1900s for a number of reasons worth recalling. One critique claims that the report "dodged the key educational questions" and "failed to deal with the crucial issues that affected the schools and would continue to affect them for decades."¹⁰ These issues included the political control over schools, the changing student demographics, ethnic tensions, wide variations in school resources, racial discrimination, and the financial constraints on schools. The failure to address these wider contextual issues of schools may have guaranteed the report's early obsolescence. In the same way, unless educational reform today addresses the full range of issues, whatever solutions we propose will be obsolete as well.

If future generations are to judge us favorably on issues of equity and excellence, 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 stringent pull toward equality can result in a tragic tumbling of all to distressingly low levels of achievement. On the other hand, we know only too well that exclusive emphasis on high standards can result in devastatingly unequal opportunities for education because of tracking and other discriminatory practices.

In addition, we must decide to what degree assessments will serve instruction as opposed to selection or accountability. Can instruments designed for administrative purposes be adapted to pedagogical ones? What do we need to know as we follow the apparent trend toward

more standards-driven and curriculum-based tests and away from the measurement of general reasoning skills? Can you tap cognitive skills more effectively via content-based tests? How will these changes interact with efforts toward national standards? And how will these changes be influenced by current national economic goals, varied business interests, the "politics of knowledge," and simple struggles over institutional "turf?"

No single innovation, whether in testing or in standard setting, will spawn the reforms equity demands, nor answer for us what purposes we have chosen for schooling. However, without a doubt, we must try to ensure that the future instruments for admission testing include the best of both multiple-choice and performance-based assessments. These new instruments must reflect high standards that can be meaningfully equated, draw heavily on new theories of cognition and new technologies, and be financially affordable. Only in this way can we be certain that *all* students in America really do have the same educational opportunity. Yet it is clear that we need to use different assessments for different purposes. Tests for purposes of accountability ought to be sample-based, as in the National Assessment of Educational Progress (NAEP), while tests for improved instruction can be designed for individual students and classrooms (for example, portfolios). Admissions tests or high-stakes tests for graduation can be complementary to both, but distinguished by high validity, reliability, and fairness. In short, admissions tests need predictive power.

Finally, equity and excellence demand a renewed collaborative effort among educational institutions across all levels, moving beyond simply greater inclusion of our national diversity and toward reimagining a wider public participation in education's purposes. Only then can we be sure our excellence will be equitable and our equity excellent. Equity, by way of rich and sustained collaboration among the members of the educational community, may be our only guarantee of excellence. As Jane Addams observed over ninety years ago:

We have learned to say that the good must be extended to all of society before it can be held secure by any one person or any one class; but we have not yet learned to add to that statement, that unless all men and all classes contributed to a good, we cannot even be sure that it is worth having.¹¹

We wish to express our deepest appreciation to Alan Heaps for his support and outstanding collaboration throughout this project. Special thanks also go to Howard Emerson and James Lichtenberg.

NOTES

1. U.S. Congress, Office of Technology Assessment, *Testing in American Schools: Asking the Right Questions*, OTA-SET 519 (Washington, D.C.: U.S. Government Printing Office, 1992), p. 103.
2. B. Alden Thresher, "Frozen Assumptions in Admissions," in *College Admissions Policies for the 1970s*, edited by C. Vroman et al. Papers delivered at the colloquium on college admissions policies, Interlochen, Mich. (New York: College Entrance Examination Board, 1978), p. 11.
3. Edward L. Thorndike, "Measurement in Education," in *Intelligence Tests and Their Uses*, Twenty-first Yearbook of the National Society for the Study of Education, edited by Guy M. Whipple (Bloomington, Ill.: Public School Publishing Company, 1923), p. 9.
4. Advisory Panel on the Scholastic Aptitude Test Score Decline, *On Further Examination: Report of the Advisory Panel on the Scholastic Aptitude Test Score Decline* (New York: College Entrance Examination Board, 1977).
5. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, D.C.: U.S. Department of Education, 1983).
6. Sol H. Pelavin and Michael Kane, *Changing the Odds: Factors Increasing Access to College* (New York: College Entrance Examination Board, 1990).
7. Daniel Koretz et al., *Can Portfolios Assess Student Performance and Influence Instruction? The 1991-92 Vermont Experience*, RP-259 (Santa Monica, Cal.: RAND, 1992); idem, *Interim Report: The Reliability of Vermont Portfolio Scores in the 1992-93 School Year*, RP-260 (Santa Monica, Cal.: RAND, 1994).
8. Stephen B. Dunbar, Daniel M. Koretz, and H. D. Hoover, "Quality Control in the Development and Use of Performance Assessments," *Applied Measurement in Education* 4, no. 4 (1991): 301.
9. National Education Association, *Report of the Committee of Ten on Secondary School Studies* (New York: American Book Company, 1894), p. 33.
10. Theodore Sizer, *Secondary Schools at the Turn of the Century* (New Haven, Conn.: Yale University Press, 1964), p. 144.
11. Ellen Condliffe Lagemann, ed., *Jane Addams on Education* (New York: Teachers College Press, 1985), p. 119.