DOCUMENT RESUME

ED 322 283 UD 027 693

AUTHOR	Ascher, Carol
TITLE	Testing Students in Urban Schools: Current Problems and New Directions. Urban Diversity Series No. 100.
INSTITUTION	Columbia Univ., New York, N.Y. Inst. for Urban and Minority Education.; ERIC Clearinghouse on Urban Education, New York, N.Y.
SPONS AGENCY	National Commission on Testing and Public Policy.; Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Mar 90
CONTRACT	RI88062013
NOTE	48p.
AVAILABLE FROM	ERIC Clearinghouse on Urban Education, Box 40, Teachers College, Columbia University, New York, NY 10027.
PUB TYPE	Information Analyses - ERIC Information Analysis Products (071) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Curriculum Development; *Educational Testing; Elementary Secondary Education; Literature Reviews; Politics of Education; *Standardized Tests; *Testing Problems; *Test Use; Thinking Skills; *Urban Education
IDENTIFIERS	Dynamic Assessment; *Policy Implications

ABSTRACT

This review of the literature on testing urban students indicates that standardized tests may not reflect accurately the ability and achievement of poor minority children. Further, new research in cognition makes clear that both teaching and testing could be structured to better prepare students for the complex thinking required by life. Since current political trends make it unlikely that the power of testing will decline nationally, or that testing will cease to drive instruction, it is crucial to reformulate assessments so that they can help alter schooling in ways that will better educate individual students to meet both their personal needs and those of society. Because short answer tests have been so important in driving learning in urban schools, and because the size of urban school systems encourages bureaucratic forms of accountability, it will be difficult to create forms of change that demand greater flexibility. However, new performance-based assessment practices offer particular hope to urban students whose gifts and needs are diverse, and who have suffered the most under traditional teaching and testing methods. Portfolios, work station assessments, certain computer-based assessments, and the variety of reciprocal teaching methods that rely on dynamic assessment all offer directions for improving urban education. A list of 66 references is appended. (MW)

Testing Students in Urban Schools: Current Problems and New Directions

Carol Ascher

ERIC®

Clearinghouse on Urban Education

Testing Students in Urban Schools: Current Problems and New Directions

Carol Ascher

TESTING STUDENTS
IN URBAN SCHOOLS:
CURRENT PROBLEMS AND NEW DIRECTIONS

Carol Ascher
Teachers College, Columbia University

Urban Diversity Series No. 100

ERIC Clearinghouse on Urban Education
Institute for Urban and Minority Education

March 1990

ERIC Clearinghouse on Urban Education
Box 40, Teachers College
Columbia University
New York, New York 10027
212/678-3433

Director: Erwin Flaxman
Assistant Director: Robin Johnson Utsey
Managing Editor: Wendy Schwartz

This publication was produced by the ERIC Clearinghouse on Urban Education with funding from the Office of Educational Research and Improvement (OERI), United States Department of Education, under contract number R188062013, and from Teachers College, Columbia University. Funds for its development were received from the National Commission on Testing and Public Policy (NCTPP). The opinions expressed in this publication do not necessarily reflect the position or policies of OERI, the Department of Education, or the NCTPP.

Copies are available from the ERIC Clearinghouse on Urban Education.

Copies are also available from the ERIC Document Reproduction Service (EDRS), 3900 Wheeler Avenue, Alexandria, VA 22304, 1-800-227-3742, both on microfiche and paper. Contact the Clearinghouse or EDRS for full ordering information.

ACKNOWLEDGEMENTS

The author thanks Linda Wing at the National Commission on Testing and Public Policy for her support throughout the preparation of this paper. Gerald Bracey of the Cherry Hill School District, and A.J. Nitko at the University of Pittsburgh, read an earlier version of the paper, and provided detailed and invaluable criticisms.

TABLE OF CONTENTS

INTRODUCTION i

A HISTORICAL PERSPECTIVE i

THE TESTING CONTROVERSY ii

CURRICULUM AND TESTING iii

INSTRUCTIONAL STRATEGIES AND TESTING iv

PREVIEW OF THIS PAPER v

URBAN EDUCATION AND THE STATE ROLE IN THE NATIONAL STANDARDS MOVEMENT 1

ACCOUNTABILITY 2

HIGH STAKES TESTING 4

SUMMARY: SOME RESULTS OF THE TESTING MOVEMENT 7

THE EFFECTS OF STANDARDIZED TESTS ON LOW-INCOME URBAN STUDENTS 9

TESTING AND ADMINISTRATIVE DECISIONS 9

THE PYGMALION EFFECT 15

TESTING AND INSTRUCTION 16

TESTING AND TEACHER PROFESSIONALISM 20

SUMMARY: THE OVERALL IMPACT OF HIGH STAKES TESTS 22

USING ASSESSMENT TO IMPROVE SCHOOLING 24

TEACHING AND TESTING FOR HIGHER-ORDER SKILLS 25

USING COMPUTERS FOR TEACHING AND TESTING 27

PERFORMANCE-BASED ASSESSMENT 28

DYNAMIC ASSESSMENT 29

CONCLUSIONS 31

REFERENCES 32

INTRODUCTION

A Historical Perspective

Since the turn of the century, educational tests developed outside the classroom by testing specialists in private companies have increasingly supplemented those devised and administered by teachers.

While teachers' tests have been thought to insure a close fit between the material assessed and the material taught in the classroom, they were also believed to contain several weaknesses. They could reflect the personal prejudices of individual teachers and, thus, work against some students. Moreover, since classrooms and schools are very different, the meaning of a grade could be difficult to determine.

The growth of commercially produced tests has a number of causes besides the desire to overcome possible shortcomings of teachers' tests. Not the least of which are the long-standing interest and ingenuity of private testing companies in finding markets, and the needs of schools for a standardized and "objective" means of sorting students. More recently, the development of electronic testing technology has both facilitated and shaped the growth of testing. Finally, the great increase in the power and prevalence of educational testing during the last ten years would not have occurred without the pedagogical and financial troubles of urban schools (G. Madaus, personal communication). Standardized tests have been seen as a way to find out what urban students were "really learning," at the same time as test scores were believed to encourage teachers and students to improve learning—with or without additional resources.

The early 1980s were marked by reform commissions whose agendas are still being followed today. In the most widely publicized educational reform report, *A Nation at Risk*, the National Commission on Excellence in Education (1983) used five indicators of educational mediocrity: results from the National Assessment of Educational Progress (NAEP), the SAT score decline, studies in functional literacy, data from the International Assessment of Educational Achievement (IEA), and test scores from the Defense Department. The Commission recommended that standardized achievement tests be administered at major transition points from one level of schooling to another "as part of a nationwide (but not federal) system of state and local standardized tests" (p. 28).

Not surprisingly, the general public caught the testing fever. A recent survey suggests that three-fourths of the public favor testing children for promotion to the next grade, the same percentage favor national achievement testing of students in their local schools to compare with students in other communities, and 89 percent favor competency testing for entering teachers (U.S. Department of Education, 1987).

Yet many educators and testing experts now believe that "our society has embraced the formal testing mode to an excessive degree" (Gardner, 1988, p. 2), and that other forms of accountability could be better used for school improvement (Darling-Hammond, 1990). The disillusionment with standardized tests has several important sources.

The Testing Controversy

There are those who argue that massive testing, like the standards movement that prompted it, has created new hurdles for just those students who were educationally most fragile (Duran, 1988; Valdes & Figueroa, 1989). Others point out that test scores have been used as a bureaucratic cover for prejudice, obscuring a societal reluctance to teach poor or minority children, since the tests "show" in advance that these students would not be able to master the more complicated material that might otherwise be taught (Neill & Medina, 1989). Although minimum competency tests may ensure that across the country graduation from high school implies the acquisition of certain skills, these tests (and the curricula they provoke) often deflect resources away from the students who are educationally most at risk (Massachusetts Advocacy Center, 1990).

Even when standardized tests are shown not to be psychometrically biased against low-income minority students—and there is still wide disagreement about whether or not they are so biased—any instrument is necessarily skewed toward one kind (or a few kinds) of intellectual and cognitive styles, and is more friendly toward students who have these styles (Anastasi, 1976; Duran, 1988; Gardner, 1988; Neill & Medina, 1989). Moreover, the very cognitive and intellectual styles assessed by current tests appear to be no better than grades at predicting college performance, and they are indifferent predictors of success in life outside of school (Gardner, 1988).

CURRICULUM AND TESTING

Another criticism raised with increasing frequency is that standardized tests, because of what they assess, have created their own hierarchy of academic disciplines. According to this criticism, the too common view has become not only that if something is important it ought to be tested; but if it can't be tested, it probably isn't important to learn. Thus, while English and mathematics have soared in educational importance, there has been a diminution of attention to laboratory science, and a near neglect of the arts (Gardner, 1988; Raizen, Baron, Champagne, Haertel, Mullis, & Oakes, 1989).

Finally, critics suggest that the short answer and multiple choice formats frequently used for testing in subjects like English and mathematics (which have been made preeminent by standardized tests) corrupt the teaching and learning process. These formats focus time and attention on simpler skills that are easily tested, and away from higher order thinking and creativity—areas that America's schools need most to develop (Resnick & Resnick, 1989).

In fact, the well-structured problems presented in multiple choice and short answer tests misrepresent problem-solving as it occurs in most life situations. The stress on facts, as opposed to opinion and judgement, is a distortion of knowledge, which always combines the three. And the assumption that learning can be decontextualized and compartmentalized into separate tiny skills, detached from life, is increasingly contradicted by the research (Resnick & Resnick, 1989).

The influences of standardized testing on education have recently prompted widespread concern, in part because many believe that the educational system in the United States must change in fundamental ways that current standardized tests may actually work against. Although current educational reform demands tests to leverage a national curriculum promoting thinking skills, existing standardized tests still feature short, choppy, superficial reading; searching for information in bits; passively recognizing errors (rather than producing corrections); and filling in pre-selected "correct" responses to other people's questions. The responses must be fast and nonreflective. Judgement, interpretation, ?nd thoughtful inference are all outside test boundaries (Resnick & Resnick, 1989).

The case is well made with reading. In most states, reading curriculum, emphasizing discrete skills, lags far behind current research demonstrating that people actually learn to read in a much more integrated manner. Yet because textbook publishers look to statewide standardized tests to guide them in selecting the skills for instruction, and the developers of statewide tests base their skills selection on current instructional materials, a vicious circle has made both reading curriculum and standardized reading tests relatively impervious to change (Peters, Wixson, Valencia, & Pearson, 1989).

INSTRUCTIONAL STRATEGIES AND TESTING

Clearly, finely-tuned assessment that assists teachers in focusing on subject matter not successfully grasped by students, and pinpoints how they are most likely to learn it, is an essential instructional strategy. Similarly, good indicators of how well or poorly schools succeed with different students are extremely useful. And assessments that predict how students will perform under specific circumstances can be invaluable to school administrators or admissions offers. Few critics of current testing practices dispute the need for appropriate measures of student learning and performance. The question for many, however, is whether tests with multiple, brief items to which there are precoded "correct" answers can be improved sufficiently to make them both fair and diagnostically useful, or whether use of existing testing instruments should be restricted and alternative means of assessment considered.

In fact, dramatically different assessment practices, from reviews of portfolios of students' accumulated work to "dynamic assessments" made during the process of teaching, are now being used around the country. These methods have the advantage of being intimately linked to the kinds of learning our nation needs—at the same time as they seem particularly suited to the diverse gifts and needs of urban students (Brown, Campione, Webber, & McGilly, 1989; Duran, 1989; Resnick & Resnick, 1989).

Preview of This Paper

The following review of student testing in urban education was supported by the National Commission on Testing and Public Policy, headed by Bernard Gifford, formerly Dean at the University of California, Berkeley. It draws heavily on those Commission papers that focus on education, particularly in urban contexts. The analysis is supplemented with references to other work being done on testing and urban education.

It concentrates on the problematic effects of testing on urban schools and urban students. While the troubled state of urban education has been an important reason for the expansion of testing, the power of tests to determine all kinds of pedagogical and administrative decisions has also been greatest in urban schools.

The first chapter analyzes the ways in which the national standards movement has changed urban education. Most important, states have assumed the power to demand higher scores, often without giving impoverished urban schools additional resources for school improvement. In fact, "high stakes testing" was initiated as a boot-strapping operation; as such, it has tended to ignore the diverse needs of both urban schools and urban students.

The second chapter draws out in some detail the effects of high stakes testing on particular aspects of urban education, including administrative decisions, placement, retention and promotion, graduation rates, instruction, and teacher professionalism. In all these areas, standardized tests appear to have rigidified and narrowed educational options for teachers as well as students, and these tests have been particular limiting to the opportunities of low-income minority students who are already at risk.

The third chapter discusses new directions in assessment that offer the potential for improving urban education. Performance-based assessments offer particular hope to urban students whose gifts and needs are diverse: these tests both enable and support a creative, thinking curriculum geared to individual needs.

Finally, some conclusions are offered, based on both the fact that testing will remain an important component in the education process and encouraging experience with the use of some new and innovative assessments.

URBAN EDUCATION AND THE STATE ROLE IN THE NATIONAL STANDARDS MOVEMENT

In the 1980s, several economic and demographic factors coincided to affect educational strategies. Amidst intensified international economic competition, the United States began moving away from industrial production to a focus on its service and information sectors. Labor force specialists forecasted that this economic shift would require high school graduates to have a capacity for thinking, reasoning, judgement, and the ability to keep learning—skills once acquired by only a small percentage of graduating students. In the demographic arena, the white middle-class school population, which had always done adequately well academically, seemed to be stagnant or even declining numerically. At the same time, the low-income, urban, minority populations, whom the schools had never served sufficiently, were growing with great speed. Thus it was clear that an increasing proportion of the newly emerging jobs would have to be held by those very students currently being graduated from, or dropping out of, school with inadequate skills. Ironically, some were even being pushed out by the very mesh of curriculum and testing that was supposed to upgrade their skills. Business was already spending an estimated $30 billion annually on formal job training and retraining (National Alliance of Business, 1987), and predictions were, that as the nation transformed itself to a service and information economy, as much as $25 billion would have to be spent yearly on remedial education (Reich, 1988).

In many cities around the country, both in and outside "the rust belt," declining industry was not only leaving families unemployed and impoverished, but was also decreasing the local tax base that might once have been available for fund schooling. Further, the federal government, which had financed so many of the equity initiatives since the 1960s, was no longer volunteering to help—except as a "bully pulpit." As the states stepped into the breach to take on more responsibility for financing education, they also began to exert new kinds of control over how the money for education was spent.

Although the gap between the achievement of poor urban children and their advantaged counterparts had been declining since the early 1970s, national assessments showed that academic achievement of students attending schools in and around large cities (i.e., with populations of 200,000 or more), where a high proportion of the residents were on welfare or not regularly employed, remained substantially below that of other types of American communities (Oakes, 1987). The low achievement scores of these students placed their schools under a variety of moral and other pressures. Moreover, a number of state initiatives were geared toward placing low-income urban school districts in an particularly tight grip.

ACCOUNTABILITY

The notion of accountability is not new in education. In the 1960s and early 70s, educators and the general public often spoke about accountability, by which they usually meant that the schools should be responsible to parents and other neighborhood people—to education's clients, or those whom public schools were supposed to serve. This accountability, often called community control, could be achieved through involving parent and other community groups in anything from raising money for school activities to hiring teachers, to deciding upon curriculum.

The reasons for the decline of "community control" are complex; however, in line with the spirit of the 1980s, school accountability began to be conceived in a narrower, more direct form. Schools were now to be accountable to parents through magnet programs and other schools of "choice." That is, parents would vote with their children's enrollment on which schools were doing their job well: they would send their children to the high-quality schools, and take them out of the poor ones. At the same time, the states, as both school funders and representatives of the people, were creating new forms of legal and bureaucratic accountability. These laws and regulations were supposed to guarantee that schools operated according to certain procedures and standards. While a "uniform system of schooling" might render parents' choices less dramatic, the state would also provide indicators of school quality, which would help parents to make educated choices.

Whereas for several decades the national emphasis had been on educational inputs—creating equal educational opportunity through, say, federally funded desegregation or compensatory education programs—growing state control in a period of tight budgets rapidly shifted the emphasis to outputs, or what schools produced. Since tests are one of the few educational options that can be imposed top-down, and are an efficient way of homogenizing diverse educational settings (Baker, 1989), testing became a major way of indicating whether or not schools and school districts were following procedures and coming up to standard.

Thus, during the 1980s, standardized testing, most often initiated at the state level, became a powerful means of sending the message that schooling could be accomplished within given financial constraints. The hope at the state level was also that tests might raise inner-city educational standards by a kind of goading action—that is, test scores would be used to prompt everyone to "try harder." In addition, testing was supposed to help tighten the curriculum: "soft" topics and courses would be pared down or even eliminated as both teachers and students worked together to increase learning in those subjects mandated by the tests. Finally, testing would ensure that the states had control over education (Glass & Ellwein, 1986).

By the end of 1984, `9 states were operating at least one statewide testing program. Twenty-nine states had enacted or approved new student examination or testing initiatives, and another 13 had such initiatives under consideration (U.S. Department of Education, 1984). Three years later, 24 states had approved state-mandated graduation requirements based on competency exams (U.S. Department of Education, 1987). As Baker (1989, p.4) has noted, "testing provides a 'we mean business' lever on efficiency and promises to demonstrate how schools can be made more efficient." Testing seemed to offer a way of bootstrapping student achievement—that is, as a method of doing more with less.

What is being tested in this profusion of mandated testing? According to a 1985 survey, in almost every case, reading and mathematics are being tested. More than half the mandated tests also test in at least one additional area: social studies, science, or language arts. Only about of a third of the states include "higher-order" questions in their testing programs. California, Connecticut, and Minnesota districts are now using writing samples to test students' writing skills (OERI State Accountability St'idy Group, 1987).

Not surprisingly, testing has not been an inexpensive venture. Most recent figures suggest that annual direct costs for developing and scoring tests now administered under local and state mandates run between $70 and $107 million. The costs are much higher when indirect state costs (still paid for by taxes) associated with test preparation and administration are taken into account (National Commission on Testing and Public Policy, 1990). For example, Texas's mandated program for testing administrators and teachers, initially budgeted at $3 million, is estimated to have cost the state $35.5 million in tax dollars when such indirect costs as in-service pay for teachers' study time and the costs of test sites were taken into account. (In addition, $42.4 million were spent by teachers and administrators who took workshops, bought preparation materials, paid for score reports, and took off time for preparation [Madaus & Haney, in press].) Based on a variety of figures, the best existing estimate of the direct cost to taxpayers of state and local testing programs, plus indirect teacher costs, is in the range of $725-915 million annually (National Commission on Testing and Public Policy, 1990).

One might ask if some of that money might have been better spent on improving curriculum and other resources of students.

HIGH STAKES TESTING

At both the state and federal levels, there has been a new and aggressive stance toward the role of testing in education. Whereas traditionally tests were likened to a thermometer which unobtrusively measures the temperature of a room, now the intrusive aspect of tests is to be acknowledged and even capitalized on. Administrators, teachers, and students alike are to become accountable for improving students' scores.

In high stakes testing, standardized test are used to bolster such important decisions as the promotion and graduation of students, the evaluation of teachers and administrators, and the allocation of funds to schools (Madaus, 1989). These tests make schools and school districts "accountable" to the states that mandate them, and increase the authority of State Education Departments over schools and school districts.

Tests as Incentives. The belief that test scores can drive school improvement has also resulted in state-based "merit" programs and other competitive incentive strategies. A number of states now reward schools or school districts for outstanding or increased achievement in test scores (as well as decreases in dropout rates, better student and teacher attendance, etc.), often with a per pupil based stipend. For example, public schools in California that test at least 93 percent of their twelfth grade students with the statewide achievement tests are eligible for an incentive award of up to $400 per student, proportional to the amount of improvement students have made (U.S. Department of Education, 1987).

Although the focus on improvement, rather than on a specific score, is supposed to take into consideration abiding socioeconomic and racial differences in test performance, there are no data yet to suggest that low-income urban schools are receiving an equitable proportion of these incentive awards. Moreover, the problem of comparing schools with vastly different financial resources and student bodies is not completely solved by the focus on improvement. First, student mobility in poor urban school districts tends to render any year-to-year comparison somewhat meaningless. Second, if education is to remain "the great equalizer," it is dangerous to institutionalize lags in the performance of low-income students by measuring change instead of achievement. Third, awarding money on the basis of test scores tends to encourage schools to eliminate low-scoring students from the test pool and so can lead to placing students in special, non-tested classes, or even to pushing them out of school (Madaus, 1989).

Competition for High Scores. In a period of industrial and other shutdowns, and heavy unemployment, the pressure states have felt to attract new business by making their educational systems look more effective has led to the "Lake Woebegone Effect," in which 50 percent of all states who administered norm-referenced tests found their students to be "above average" in English, and 75 percent found their students to be "above average" in mathematics (Glass & Ellwein, 1986). Such a statistical anomaly has two causes: first, any test that drives curriculum will for a time result in increased test scores, which means that most states have produced narrowly positive results by their focus on accountability testing. Second, although standardized tests have been renormed in the 1980s to account for raised scores, some states have understandably discovered that it is easier to produce good news if they continue to use the 1970s-normed tests, which testing companies still sell.

There has also been a drive to create a nationally competitive system of state ranking. In 1984, the U.S. Department of Education began publishing its "Wall

Charts" on which states were ranked according to standardized achievement scores and other variables. In March 1987, after years of resistance to state-by-state comparisons, the Education Commission of the States recommended that the National Assessment of Educational Progress be expanded to permit such comparisons. And the September 1989 Governor's Conference resulted in a consensus to move toward both a national curriculum and national test scores.

Some states are also initiating new testing programs aimed at enticing corporations to their state. Michigan, for example, is pioneering a controversial "employability" test, and Hawaii is following close on its heels.

School Finance. Unfortunately, state pressure to raise test scores has rarely been accompanied by general funding to improve impoverished urban schools. As of 1986, only ten states had initiated programs to improve the educational outcomes of at-risk students (Oakes, 1987). Yet, correlations of SAT and ACT scores with the level of funding of school districts suggest that students in poorly funded schools (who are typically low-income minorities) perform less well on standardized tests (Harris, in Sosa, 1988). In fact, a number of structural features in state aid systems have worked against urban schools with large poor and minority populations in this retrenchment period. For example, Average Daily Attendance formulas, which form the bases for state aid, tend to discriminate against urban school districts with high absentee rates. Also, state dollars have been more likely to go to "excellence" projects than to support for the education of disadvantaged students (Ascher, 1989). Yet the pressure has remained—and even grown—to improve the educational preparation of the next generation of workers, which, in fact, means improving the success of urban schools.

The use of state regulatory power to improve schooling has raised concern among educators, both in and outside urban systems. They suggest that state efforts should be directed at ensuring equal inputs, rather than legislating program and standards. As Wise and Gendler note, when the state regulates outputs, its "effort to produce equal education ends up degrading learning for all. Individuality, creativity, and depth are lost; all that is retained is uniformity, conventionality, and trivial skills" (1989, p. 36). By contrast, when a state regulates inputs, it encourages local initiative, equalizes the capacity of poor districts to secure a sufficient and highly qualified teaching force, and permits schools from poor districts to choose among curriculum and equipment options, just as wealthy districts do.

SUMMARY: SOME RESULTS OF THE TESTING MOVEMENT

Over the past decade, the pressure on the nation has been great to improve education, particularly in urban schools. In a period of tight, even declining, education budgets, the federal government and State Departments of Education have sought to positively affect student achievement by mandating standardized tests and publicizing test scores. The hope was that such "high stakes testing" would create a kind of bootstrapping action, raising educational standards at the same time as it cut away the "soft" topics not addressed by the tests.

High stakes tests have by now become ubiquitous, and they are used to bolster a range of critical pedagogical and administrative decisions. Even awards to schools are given on the basis of "merit" as judged by students' test scores. All of this has created enormous pressure to generate high scores—which schools have done, not surprisingly, both through teaching to the test and through creative scoring.

The tests have also reinforced a hierarchy of academic subjects: they show that mathematics and reading are valued (although real reading is, in fact, generally poorly tested) more than social studies or art. And they have rewarded teaching that results in correct responses to multiple choice questions, rather than in the ability to write thoughtful or creative essays. In fact, several testing experts suggest that these high stakes tests may be more important in determining what and how schools teach than in measuring how smart American students are—or even what they are learning (Haertel, 1989; Haney, 1989). As Haertel points out, this symbolic use of testing "is not benign. It directs attention from the root causes of low achievement, and encourages the simplistic idea that if policy makers just insist upon results, then teachers and administrators can somehow manage to provide them" (p. 40).

Ironically, massive testing has also been a costly venture. At a time when urban schools in particular have operated under terrible constraints, standardized tests have run in the hundreds of millions of dollars—far beyond what anyone ever imagined. Equally distressing, such testing has not resulted in a better educated student population, although test scores have gone up. In fact, there is some evidence that schooling has been narrowed and skills "dummed down" on the basis of the tests.

The specific effects of high stakes tests on a variety of decisions in urban schools will be analyzed in the next chapter. Here it is important to conclude with the reminder that standardized tests have directed attention away from the financial straits of urban school districts, shifting responsibility for failure from financially and intellectually impoverished systems to inadequate teachers or poor students.

THE EFFECTS OF STANDARDIZED TESTS ON LOW-INCOME URBAN STUDENTS

> In Boston...the emphasis on basic skills, high stakes standardized testing, rigid promotion and graduation standards, and the notion of student "readiness" support the sorting, labeling, and tracking of students at all levels and thus defeat any attempts to improve schools' educational mainstream (Massachusetts Advocacy Center, 1990).

While every public school in the nation now regularly uses standardized tests for accountability, the proliferation of these tests was largely spurred by the lower achievement of inner-city school students (G. Madaus, personal communication). Accountability tests are not only used more extensively in low achieving urban schools (Neill & Medina, 1989), but they appear to have a greater effect on curriculum planning, classroom assignments, and funding allocations in low SES schools (Dorr-Bremme & Herman, 1986).

According to a three-year national study by the Center for the Study of Evaluation, standardized tests influence the fate of students, teachers, and curriculum far more in schools serving low-income and minority students than in other schools. Districts serving disadvantaged students are more likely to establish school goals, and principals in schools serving these students look at tests more often than do principals in middle-class schools (Dorr-Bremme & Herman, 1986). Other researchers point out that testing systems are more powerful in schools with low-income minority students because they serve as a routine part of tracking and other bureaucratic responses to heterogeneity (Neill & Medina, 1989). And there is evidence that the curriculum in urban schools is more affected than in other schools by efforts to teach for testing (Dorr-Bremme & Herman, 1986). Finally, researchers have suggested that the attempt to increase test scores has accelerated suspension, failure, and dropout rates as schools attempt to get rid of "troublesome" students (National Coalition of Advocates for Students, 1985).

TESTING AND ADMINISTRATIVE DECISIONS

As early as the turn of the century, Alfred Binet, the father of intelligence tests, warned against the potential for abuse in mental testing. He feared that teachers would find standardized tests "an excellent opportunity for getting rid of all children who trouble us," and that there would be an element of self-fulfilling prophesy in test results, since "it is really too easy to discover signs of backwardness in an individual when one is forewarned" (Binet, 1905, p. 170; in Brown, et al., 1989).

Despite Binet's warnings, an early and continuing use of standardized testing has been to divide school populations into groups of students whose test scores suggest that they are similar in level of achievement. This use of testing to create homogeneous groupings has been particularly prevalent in cities educating students from diverse backgrounds (Neill & Medina, 1989). As early as 1925, a survey of schools in cities over 10,000, for example, showed that the primary function of all standardized achievement and ability tests was to place students in homogeneous classes (Resnick, 1980).

Massive testing also received an impetus from civil rights and other equity legislation of the 1960s and 70s, which required evaluations of programming for continued funding (Frederiksen, 1984). At the same time, however, testing, and the homogeneous groupings it facilitated, minimized the effects of school desegregation by resegregating students into largely homogeneous racial groups. Although homogeneous groupings have been determined to be pedagogically unsound (Slavin, 1980), recent surveys suggest that most urban schools still divide their students into classrooms by ability, and that these determinations of ability are made primarily on the basis of standardized test scores (Oakes, 1985). In fact, the importance of testing to instruction has reinforced the bureaucratic tendency to fill classrooms with students whose test scores are similar—and thus who can be more conveniently helped to meet the next rank on an improvement standard (Oakes, 1985).

Even when a test is only part of the protocol informing an administrative decision for any student, two issues are at stake. The first is whether the test is a good predictor of how the student will do in a particular classroom or with a particular instructional method. For example, testing and measurement specialists often suggest that the predictions made for bilingual students on the basis of standardized tests are not as accurate as those made for native English speakers, and that the tests underpredict how well these students will do in future academic settings (Valdes & Figueroa, 1989). A second concern in using a test to inform an administrative decision, particularly with low-achieving students, is whether the decision influenced has any educational imperative—that is, does the student's test score merely lead to "labeling and discarding" (Scarr, 1981, in Jones, 1988), or does it provide real information to ensure that the student will profit from the instructional placement (Mehrens, 1989).

Testing and Placement. Testing is currently used in several types of placement contexts: to determine children's "readiness" for school, to decide on their placement in achievement groupings or tracks (including special education classes), and to make decisions about whether they are eligible for compensatory education programs.

The pressure on schools to raise test scores has spawned the widespread, but controversial, practice of using standardized testing to determine entry to kindergarten or first grade. The logic often cited here is that the easiest way to raise test scores in the early elementary grades is to start with classes of children whose scores are already good (Cunningham, 1989). Even when administered individually, however, standardized tests for young children are among the least valid and reliable exams—some 35-50 percent of those children tested are misidentified as not ready to start first grade (Cunningham, 1989). Further, readiness as judged by the tests is not likely to be evenly spread over all population groups: in South Carolina, 60.9 percent of *all* black five-year-olds tested as ready for kindergarten compared to 82.5 percent of whites in 1984 (Neill & Medina, 1989; Wofford, 1990). The low validity of readiness tests is partly because of the near impossibility of administering such tests under uniform conditions, and partly because chronological age affects test scores at this age more than later, and so even a month or two can make a vast difference in test scores. And the extremely high number of black students identified as not ready is largely the result of the correlation between test scores and social class.

Whatever predictive validity a particular student's test scores has, there is no instructional validity to the placements made as a result of readiness tests. That is, there is no evidence that pre-kindergarten or first grade retention—the most common administrative decision resulting from low test performance—actually leads to higher achievement (Cunningham, 1989). Yet in Georgia, the state that pioneered in standardized early childhood placements tests, students who failed their readiness tests were put in either a "transitional kindergarten" or a "pre-kindergarten" class—neither of which had any clear pedagogical effect in leading to greater readiness in the subsequent year (Shepard, 1989).

Several of the most well-publicized legal cases against testing in relation to placement have been connected with the use of intelligence tests for diagnosing learning disabilities. This is because IQ tests have tended disproportionately to identify blacks, Hispanics, and Native Americans as "learning disabled" or "mildly mentally retarded." In some cases, Hispanic students have been so identified at three times the rate of other students (Valdes & Figueroa, 1989). The disproportionate number of blacks being funneled into "educable mentally retarded" classes in California occasioned the 1979 court case, *Larry P. v. Riles*, which placed the burden of proof on the schools to rebut a presumption of discrimination. Although a different outcome resulted from a court case the next year, educators are now aware that testing decisions easily lead to classroom placements that both stigmatize students and fail to provide them with appropriate instruction.

The use of testing in schools serving low-income students is increased by special state and federal education programs for the disadvantaged that mandate standardized tests. These are the schools that tend to have Education Consolidation and Improvement Chapter 1 programs requiring periodic testing for accountability. In fact, although current guidelines require only annual reporting on the program's impact on students, many districts test both in spring and in fall (Kennedy, Birman, & Demaline, 1986).

Recent research suggests that the disservice done to disadvantaged children by traditional IQ and achievement tests might be ameliorated by dynamic forms of assessment. This new type of testing will be discussed in the next chapter. Here it is enough to say that, in contrast to static assessment, which focuses on what children already know, dynamic assessment elicits children's capacity to learn—which is what is most important in any educational placement.

Testing and Retention/Promotion. The administrative decision to use retention, as opposed to other ways of handling students who have not mastered a year's curriculum, has waxed and waned over the past century. Current research shows that students rarely improve their achievement on the second round; the exceptions occur when retained students receive special instruction that does not merely repeat the same curriculum (Ascher, 1988b). Nevertheless, the "get tough," high standards emphasis of the 1980s has tended to pit retention against "social promotion," an alternative that has come to imply low or inadequate standards, and standardized tests are now being used as a precondition for passing through certain "gates"—often at the end of the first, third, and eighth grades.

Since minority students are more likely than whites to test at the lower end of achievement test scores (as well as to be seen as troublesome by teachers), they have retention rates three to four times higher than those of their white peers. Among blacks, males are particularly at risk for retention. In court hearings claiming prejudice in decisions to retain, the courts have generally upheld these decisions as academically-based. However, in several cases where a disproportionate number of blacks failed to perform satisfactorily on standardized tests, particularly if the school was previously segregated, the courts have asked school systems to justify their retention/promotion policies (Stroup & Zirkel, 1983, in Ascher, 1988b).

One area where retention has traditionally been thought most useful—and least likely to have harmful side effects—is in the early years. However, when decisions are made on the questionable basis of standardized achievement tests, retention in first grade appears of no significant benefit by the time the students are fourth graders. Nor does testing with such instruments as the Metropolitan Reading Test (MRT) or the Gesell Preschool and School Readiness Test, which are supposed to measure developmental age, improve retention decisions. Although it may appear benign to hold back children who are "developmentally young," in fact these tests measure the construct IQ, highly correlated with social class—which ensures that a disproportionate share of the children who score "in need of retention" will come from low socioeconomic backgrounds (Cunningham, 1989).

Evidence is also growing that students retained in the elementary grades tend to be those who drop out later. Although retention may not cause dropping out, there is a strong connection. This means that insofar as standardized tests increase the retention rate in the early grades, they may also be indirectly accountable for later dropping out. A Cincinnati Public Schooi analysis of the system's dropout data, for example, found that students with one retention had a 40-50 percent chance of dropping out of school, those with two retentions had a 60-70 percent chance, and those with three retentions rarely graduated (OERI Urban Superintendents Network, 1987, in Ascher, 1988b). Although social promotions are no solution, the fact that minorities have retention rates three to four times higher than those of whites (Ascher, 1988b) suggests that the burden of proof should be on schools to justify their use of standardized tests in their retention/promotion policies.

Testing and Graduation Rates. The pressure on schools to graduate a literate work force has led to increased use of minimum competency tests. Ironically, students who complete high school are not deficient in the basic skills tested (Frederiksen, 1984; Resnick, 1980), but these are not the skills that economic forecasters cite as most needed in our service and information society (National Commission on Excellence in Education, 1983).

While minimum competency tests are not ensuring a literate work force, they are also not improving education for at-risk students. Researchers have recently attempted to discover whether these tests are exacerbating the worrisome urban dropout rate. So far, the findings are uncertain—but they do suggest a variety of warnings. Glass and Ellwein (1986) note that when standards on tests are raised, safety nets are strung in the form of exemptions, repeated trials, softening cut-scores, tutoring for retests, and the like. When this happens, even though the dropout rate is not increased, the value of the test scores in measuring learning is obviously decreased. Caterall (1987), however, has somewhat more equivocal findings about the effects of competency tests on students' dropping out. In a review of state data collected between 1982 and 1985, he found that graduation rates were negatively correlated with having a required test for graduation and with having instituted a test. Nevertheless, Caterall also points out that schools use "tests of differing nature and lengths, differing calendars for initial testing and retesting, differing remediation programs for test failers, [and] varying numbers of retest options allowed" (p. 8). While graduation tests may prove "convenient" for those schools that wish to push certain students out, other schools may go to great lengths to encourage passage by academically marginal students.

Finally, the connection between testing and retention may also become the indirect connection between testing and dropping out. Tests given in the elementary grades, for example, may lead to retention, which may increase the likelihood of a student's dropping out eight or ten years later.

THE PYGMALION EFFECT

The argument that test scores can goad teachers and students toward greater diligence assumes a positive feedback mechanism: low scores will give schools the push to improve; high scores, the pride to do better.

Unfortunately, feedback from test scores is neither simple nor always positive. In the late 1960s, a famous study described how teachers' expectations for their students (in this case, created by fake IQ test results given by the researchers) influenced the teachers' treatment of these students, which in turn created differences in students' performance and thus in the grades the teachers gave them (Rosenthal & Jacobson, 1968). *Pygmalion in the Classroom* became quite controversial and has been replicated many times with somewhat varied results. However, Jussim (1986) used related research to create a psychological and social model of three processes—teacher expectation, differential treatment, and students' psychological and behavioral reactions—which argue convincingly for the Pygmalion dynamic. In another attempt to understand the Pygmalion effect, a meta-analysis of research on the effect by Raudenbush (1984) suggests that the most important variable in determining whether teachers are influenced by test information is the amount of their prior contact with the students.

Madaus' study (1989) of a low stakes testing situation in Ireland creates a further wrinkle in the Pygmalion effect, at the same time as it affirms the phenomenon. His Irish data indicate that test information and teachers' perceptions did not interact the same way for boys and for girls. Teachers who received test information tended to be more influenced by it when dealing with their male students. At the same time, the boys' subsequent performance did not appear to align to meet the teachers' changed perceptions. On the other hand, while test scores did not influence teachers to alter their initial perceptions of their female students, girls' performance tended to align with the teachers' original views. Madaus also found evidence that test information had a stronger impact on expectations in urban than in town or rural schools. Finally, "while test information tended to benefit pupils whose test scores were higher than their teachers' assessments of their ability or achievement, it tended to work to the detriment of pupils for whom the teacher's original assessment was more favorable than the test results" (p. 76). Madaus also notes that all these effects might well be exacerbated in a high stakes testing situation.

Test scores also affect students' views of themselves—even without feedback from teachers. Research on SAT score feedback, for example, suggests that students take these scores more seriously than the grades they receive from teachers. Moreover, these standardized test scores influence their choice of colleges. For women and minorities, whose scores can under-predict first year college grades, academic self-perceptions may be set too low and the students may not apply to academically demanding colleges for which they are, in fact, qualified (Teitelbaum, 1989).

Testing and Instruction

When researchers speak of tests as biased against urban schools or low-income minority students, they are often speaking as much about the inequities in student preparation as about technical biases in the tests. In fact, many minority advocates look to test scores, however inadequate, as the "bad news" they must attend to if they want to leverage school improvement (Duran, 1988; Jones, 1988).

Student Preparation. Data from a variety of national testing programs suggest a strong relationship between academic preparation and test scores. Analyses of SAT scores, for example, show that for all groups there is a corresponding increase in verbal and mathematics test scores with each increase in the number of courses taken in high school. Unfortunately, this unsurprising relationship between serious preparation and test scores places many minority students at a disadvantage. According to High School and Beyond data, only 26 percent of Hispanic high school seniors, and 32 percent of black high school seniors, follow a college preparatory curriculum—compared to 39 percent of whites. And Hispanic students are overrepresented in high schools with fewer resources and in curriculum tracks within high schools that have less demanding courses (Duran, 1988; Pennock-Roman, 1988).

There are few studies of courses taken by Native Americans, but one study of four Bureau of Indian Affairs boarding schools found that only one had a first-year algebra, and none offered advanced algebra, trigonometry, geometry, or calculus (Chavers & Locke, 1989).

There is also evidence that students in segregated schools are at a disadvantage when taking tests—probably because these schools tend to have poorer academic resources. Data from the California Assessment Program (CAP), for example, suggest that Hispanic students who attended the most segregated schools were the most likely to earn lower CAP reading scores (Duran, 1988).

Testing and a Measurement-Driven Curriculum. Although the current national and state involvement in test scores offers little help in planning the best learning strategy for individual students, the rewards and sanctions of test scores for administrators, teachers, and students are now huge. The pressure to show good—and regularly improved—scores on short answer multiple choice tests has had a number of consequences for individual schools and school districts. One has been the tendency to ignore higher-order skills, since multiple choice and short answer items are geared to elicit facts, not show evidence of complex cognitive processes. Another consequence has been a significant increase in time spent preparing students for tests, especially at the elementary level. The increase is especially notable in low-income schools where drilling was always a more prevalent form of instruction. This curriculum shift is significant, to the point that "testing is driving the curriculum in economically disadvantaged areas to a greater extent than elsewhere, particularly in elementary schools" (Dorr-Bremme & Herman, 1986, p. 75). A 1988 study identified 12 commercially prepared achievement test preparation programs, including two film strips by the Educational Testing Service, that feature drill and practice ditto masters and workbooks presenting students with multiple choice questions. These commercial materials increase the proportion of time that children spend on short answer drill work at the expense of the higher-order thinking skills that the nation supposedly demands (Madaus & Haney, in press).

Advocates of measurement-driven instruction see the influence of testing on instruction as positive, since, in their view, tests establish clear performance standards and clarify for students and teachers what they are expected to do, while allowing teachers the latitude to attain those objectives as their professional judgement dictates (Vickery, 1988). However, experience with state-mandated tests suggests that teachers do not quickly accomplish their test-preparation objectives and then go onto other learning activities. Thus, even when teachers and students are not directly preparing for tests, the increase in testing has reshaped instruction to a measurement-driven curriculum, largely devoted to drill and practice. As Haertel (1989) points out, tests have revived the "factory model" of education by routinizing and standardizing learning to "testable student outcomes as its sole, dreary product."

Whether one speaks politely of "curriculum alignment" or, more critically, of a "factory model" of instruction, evidence suggests that nationally there has already been a decrease in instruction in science, writing, problem-solving, and analytical reasoning (Cunningham, 1989; Resnick & Resnick, 1989). In fact, the trickle-down of testing to the early elementary years has also created a test-driven curriculum in kindergarten and the first grade, where students must be taught quantifiable skills (math and reading), often to the exclusion of social skills or even a deeper development of their cognition (Cunningham, 1989). In Texas, where materials suggesting instructional strategies for each of the objectives on the state assessment were provided to teachers, a classroom study showed that teachers increasingly aligned both the range of skills taught and the form of these skills to the test (Shepard, 1988). Not surprisingly, a recent survey of early childhood educators revealed that 60 percent believed they were being forced to teach in ways that were harmful to their children as a result of pressure from year-end standardized tests (Shepard, 1989).

Test-Taking Skills vs. Conceptual Understanding. One of the reasons suggested for the lower test scores of black and Hispanic students is that they suffer from inadequate test-taking skills—among them, "psyching out" the test, or guessing, are of obvious importance. However, the experts divide on whether one should work directly to improve test-taking skills. This may be because different studies on the effects of coaching show rather varied results, with some indicating that low-achieving minorities can gain substantially—even more than whites—and others showing extremely modest gains (Bond, 1988). Evidence also suggests that instruction aimed at conceptual understanding may enable students to do better than drilling on test-taking skills directly (Carpenter, Fennema, Peterson, Chiang, & Loef, 1988).

Diagnostic research on individual students has also attempted to isolate problems in reasoning that may be associated with low scores—and these can be corrected. For example, one study shows that the poor scores of black high school students on standardized measures of quantitative reasoning may be the result of too much time spent on solving routine sub-problems that could be speedily resolved if their knowledge were automatic, (Bond, 1988). Bond's study suggests that the poor performance of the black students is due largely to "inert, unintegrated knowledge" and inefficient problem-solving procedures that approximate a random search; and that one reason for these poor skills is that mathematics as currently taught does not stress real problem-solving. According to Bond, high school mathematics "tends to emphasize the acquisition of declarative knowledge necessary for problem-solving but...the next pedagogical step (systematic and sustained instruction in using that knowledge to solve problems) has either not been attempted or, if attempted, has been unsuccessful" (p. 12). In fact, Bond's research yields just the kind of diagnostic feedback that all tests should offer. At the same time, it suggests that blacks and Hispanics are being tested on material they may not have been taught as adequately as others—or even at all.

The discrepancy between what the tests assess and what current theory suggests about learning has also been felt in the field of reading, where it

> creates an ironic situation for policy makers who hope to use tests as a lever to improve education. They want high quality education, but when they use outmoded reading tests....[i]t is unlikely that such efforts will improve the quality of educational practice, because tests will drive the curriculum in an inappropriate direction (Peters, Wixson, Valencia, & Pearson, 1989, p. 3).

The awareness of the problems inherent in current testing instruments for reading has generated some interesting efforts to alter what the tests test. The experiments in Illinois and Michigan, for instance, aim to move students toward the ability to comprehend, interpret, synthesize, evaluate, and draw inferences from written material (Peters, et al., 1989).

TESTING AND TEACHER PROFESSIONALISM

> One of the most attractive aspects of professional work is the way professionals are treated in the workplace. Profssionals are presumed to know what they are doing, and are paid to exercise their judgement. Schools, on the other hand, operate as if consultants, school district experts, textbook authors, trainers and distant officials possess more relevant expertise than the teachers in the schools. Bureaucratic management of schools proceeds from the view that teachers lack the talent and motivation to think for themselves (Carnegie Forum on Education and the Economy, 1986, p. 57-58).

State pressure to improve test scores has led to district policies that link effective teaching with specific (often rote) curriculum and testing programs. Some districts also provide inservice training for teachers in using the curriculum and testing method. San Diego's Achievement Goals Program includes district-based inservice training in mastery-learning techniques, and Pittsburgh's PRISM is closely connected to the Monitoring Achievement in Pittsburgh (MAP) program (Oakes, 1987).

Yet the growth of testing has generated concern about the loss of autonomy and professionalism among teachers. This has been heard from teachers themselves, who complain that tests are used too heavily to reach important decisions and that the standardization of teaching imposes constraints on their ability to meet students' needs. In fact, there is evidence that "the more impact standardized tests have on instruction, the more teachers resist their use" (Darling-Hammond & Wise, 1985, p. 321). Almost half the teachers in one survey reported that their morale is worse and that political interference and paperwork has increased (Carnegie Foundation for the Advancement of Teaching, 1988). Interestingly, most teachers tend to believe that other teachers' autonomy has been more threatened by testing than has been their own. While 60 percent of all teachers in one sample reported that the increased emphasis on standardized testing had affected their own curriculum, they believed that it had affected the curriculum of 95 percent of all other teachers (Darling-Hammond & Wise, 1985).

In fact, curriculum innovation has been an important area of professionalism to suffer from state-mandated tests. As tests increasingly drive curriculum, teachers have been forced either to create a dual system of teaching and testing—one for what they consider good pedagogy, and the other for test score improvement—or, more likely, to give up on their own instructional theories and plans. As one testing expert put it, "Once the curriculum has been reduced to a string of tiny hurdles, all equally important, it becomes difficult to experiment with alternative conceptions of content, organization, or instructional approach" (Haertel, 1989, p. 38).

Researchers and social critics point out that loss of autonomy and professionalism is as much a threat to the morale of existing teachers as it is to for the recruitment of good new teachers. This is especially a problem in urban schools, where the teaching staff is aging, and where great shortages are already evident, particularly among minority teachers. As the Carnegie Forum notes, "If the schools are to compete successfully with medicine, architecture, and accounting for staff, then teachers will have to have comparable authority in making the key decisions about the services they render" (1986, p. 58).

In fact, those schools where teachers have most severely felt the loss of autonomy over both students' progress and curriculum are the schools serving low-income students (Dorr-Bremme & Herman, 1986). It is in these schools that the state-mandated tests also tend to show low scores. Thus the combination of tested low performance scores and loss of professional autonomy may be an under-discussed aspect of the low morale of teachers in low socioeconomic schools. It may also be one reason why seasoned teachers try to get transferred out of urban schools.

Finally, it is clear that innovations in testing and instruction cannot be thought of as isolated from the issue of teacher professionalism. Haertel (1989) and Darling-Hammond and Wise (1985), among others, have suggested that any initiative to standardize the curriculum or impose particular assessment methods should be evaluated in the light of its probable effect on teacher professionalism. And, in a particularly strong statement on the need for professional autonomy, the Carnegie Forum (1986) maintains that:

> within the context of a limited set of clear goals for students set by state and local policy makers, teachers, working together, must be free to exercise their professional judgement as to the best way to achieve these goals. This means the ability to make—or at least to strongly influence—decisions concerning such things as the materials and instructional methods to be used, the staffing structure to be employed, the organization of the school day, the assignment of students, the consultants to be used, and the allocation of resources available to the schools (p. 58).

SUMMARY: THE OVERALL IMPACT OF HIGH STAKES TESTS

The proliferation of state-mandated testing in the 1980s has catapulted the power of standardized tests to a new level. For the first time, students and teachers are being evaluated by high stakes tests. In effect, the promotion and graduation of students, the evaluation of teachers and administrators, and the allocation of resources to schools have all become dependent on standardized test results. Unfortunately, it is not at all clear that these high stakes tests are really giving educators, the business community, parents, or students the information they want about schooling.

Madaus (1989) suggests several principles that govern how high stakes tests influence schooling. Most important, the more tests are used for educational decision-making, the more likely they will distort the process they are intended to monitor. This is the power of testing over curriculum and teaching, and it implies three other principles. First the belief by teachers, students, and administrators that the tests are important causes teaching and learning to the test. (This includes teaching to the form of the questions on the test, as well as to their content.) Second, with each year of high stakes testing, a tradition of past tests develops, which eventually *de facto* defines the curriculum. Third, "when test results are the sole or even partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement" (p. 85), which, in turn, increases the power of tests to shape education.

The growth of state mandated testing programs is worrisome, especially in its effects on urban schooling. Testing has tended to reinforce the bureaucratic aspects of decision-making, which already weigh so heavily on large schools. Although tests appear to set uniform standards, creating homogeneity out of diversity, in fact, standardized tests easily reinforce traditional hierarchies of race and class. In urban schools, testing supports tracking and other homogeneous groupings, which have long been shown to hurt poor and minority students.

Inside the urban classrooms, without academic resources comparable to schools serving more affluent students, testing has too easily been a way of passing the blame for inequality in offerings and resources onto teachers and students. As important, mandated tests have tended to narrow instructional practices to a measurement-driven curriculum, or to teaching directly for improvement on test scores. They reinforce the drilling of "basic" skills, in which many urban schools were already over-invested, while ignoring the higher-order skills that enable real learning and that our society presumably needs.

Finally, testing has restricted teacher autonomy and professionalism, creating yet another reason why urban schools can be thought of as unattractive places to teach.

Because all these effects of testing programs are critical to good learning, and especially important in low-income schools, they raise issues that should be considered when testing programs are instituted.

USING ASSESSMENT TO IMPROVE SCHOOLING

> If life consisted solely of schooling, most formal tests would serve their purpose well—though last year's grades would fulfill the same predictive purposes equally well. However, schooling is supposed to be a preparation for life, and there is ample evidence that formal testing alone is an indifferent predictor for success once one has left school (Gardner, 1988, p. 10).

Current political trends, as well as the increasingly sophisticated possibilities of testing technology, make it unlikely that the United States will go back to the low stakes testing of yesteryear. Many good schools now operate with a dual system of teaching—one for the standardized tests and another for good instruction (Darling-Hammond & Wise, 1985). Obviously, this is a wasteful and exhausting enterprise for both students and teachers.

Resnick and Resnick (1989) have suggested that in any test-driven educational reform movement, curricular changes will be based on material that tests assess; what is not assessed is not likely to be taught, so assessments must be created that State Departments of Education and other test purchasers want teachers to teach to. As this observation suggests, a correlation with a desired educational outcome is not a sufficient criterion for purchasing a test. ETS, for example, claims that the multiple choice questions on the English part of its exam correlate highly with good writing skills. Even if this is so, the fact is that writing instruction has declined as a result of the multiple choice form of the test: teachers won't teach writing until tne tests use writing as the *form* and *content* of the test (Resnick & Resnick, 1989).

Similarly, if educators want to assess higher-order thinking, they cannot use short answer exams that encourage superficial knowledge, quick responses, and a sense that there already is a "correct respcnse." They must stop decontextualizing and begin instead to test thinking in action, and to evaluate learning through station activities, portfolios, and other "performance" devices that show the complexity of students' achievement. This may entail a dramatic break with the past, and it seems a tall order to administrators already overburdened by problematic schools, but the alternative is simply to proceed further down a wrong path.

Moreover, a number of school programs around the country are already focused on new types of testing. For example, Arts Propel in Pittsburgh reviews projects, portfolios, and reflective interviews (Wolf, 1988; Gardner, 1989); and Project Zero in Cambridge assesses portfolios (Brandt, 1988). These open-ended assessments provide insight into students' abilities to ask and pursue worthwhile questions, show the range of processes students command, and sharpen students' capacities to reflect on their own work. The scoring of these projects is still in early stages; however, the open-ended writing assignments currently being tried in a number of states show that it is possible to score open-ended questions with high reliability (Resnick, 1989). Finally, the use of portfolios in England and Wales as part of a highly structured and rigorous national assessment system (Goldstein & Wolf, 1989) provides important direction for those attempting to develop more complex assessments for use on a mass scale.

TEACHING AND TESTING FOR HIGHER-ORDER SKILLS

Findings from cognitive research provide new directions for teaching and testing for higher-order thinking. In particular, they suggest new, more intimate links between teaching and testing that connect both more directly with real life situations.

First, cognitive research suggests that there is a wide range of human intelligences, as well as cognitive styles. Therefore, it is unlikely that any uniformly presented curriculum will suit everyone equally, or that everyone will learn from it at the same rate. Phrased more positively, it means that students will learn best when their strengths and weaknesses are identified early and they are presented with curriculum tailored to their needs and interests (Gardner, 1988).

Second, cognition and learning specialists are refocusing their conception of intelligence to "the ability to learn." Thus, new assessments test intelligence with tasks that cause the test-taker to learn in the process of taking the test. The students' intelligence can be assessed by how far he or she can proceed with incomplete instructions, or by how many further instructions must be given along the way—that is, by a process called dynamic assessment (Sternberg, 1989; Brown, et al., 1989).

Equally important, intelligence in this view is a dynamic, rather than a static, quality. It can be improved through instruction—that is, through learning how to learn. In fa^ct, one of the critical findings in this area is that traditional tests, which provide a static picture of what students know at one point, rather than a measure of their capacity to learn, are especially poor predictors of academic achievement for disadvantaged children, because they gravely underestimate the capacity of disadvantaged children to learn. "If the interest is in predicting the learning trajectory of different students, the best indicant is not their IQ or how much they know originally, nor even how readily they acquire new procedures, but how well they understand and make flexible use of those procedures in the service of solving novel problems" (Brown, et al., 1989, p. 63).

Third, studies point to the different rates at which learning takes place at different periods in a person's life, as well as to the individual variations in these rates. The idea that everyone ought to have mastered a particular amount of curriculum by October, by January, and by May, contradicts what is now known of the human mind, which appears to work in periods of rapid accumulation followed by hiatuses. Related to this, it is not axiomatic that simply adding hours or days to the amount of time spent in school will always increase learning (Heyns, 1978; Ascher, 1988a).

Fourth, careful analyses of learning make clear that one cannot teach reading and mathematics as isolated mechanical skills and hope to add on a thinking curriculum, if time permits. This is because all genuine learning of "the basics" contains higher-order thinking. Moreover, there is no clear distinction between knowledge and thinking. All knowledge involves reasoning, judgement, and opinion. Similarly, an individual's reasoning and judgement develop hand in hand with his or her growth of knowledge in a particular area. In fact, even aptitude is best elicited as a student learns a subject—not before (Gardner, 1988; Resnick & Resnick, 1989).

Fifth, a range of research suggests that even experts often fail on "formal" measures of their calculating or reasoning capacities, although they excel at precisely these same skills in their ordinary work. Rephrased in the language of current tests, abilities are not always either decomposable or decontextualizable, and learning is best assessed in context (Gardner, 1988; Resnick & Resnick, 1989).

Sixth, investigations into thinking show that cognition does not take place "inside" a person's mind, isolated from the surrounding world. Instead, cognition is better thought of as a three-way *intersection*: the individual with his or her skill, knowledge, and aims; the structure or domain of knowledge he or she confronts; and the role he or she is being asked to play in the institution.

> Viewed more broadly, it makes sense to think of human cognitive competence as an emerging capacity, one likely to be manifest at the intersection of three different constituents: the "individual," with his or her skills, knowledge, and aims; the structure of a "domain of knowledge," within which these skills can be aroused; and a set of institutions and roles—a surrounding "field"—which judges when a particular performance is acceptable and when it fails to meet specifications (Csikszentmihalyi, 1988, in Gardner, 1988, p. 13).

Using Computers for Teaching and Testing

One of the most likely areas of teaching and testing to develop from the new, more complex cognitive model is computers. This is because computers are such a convenient tool for creating instructional models that are simultaneously testing mechanisms. They offer simulations of real life problems. They can provide prompts to individual step-by-step learning, adjust tasks, and log responses—at the same time as each interaction between student and computer becomes its own information for assessment. Computers also enable classroom teachers to work with individualized programs that are part of larger district, state, or even national testing and learning programs (Raizen, et al., 1989).

On the other hand, several researchers suggest good reasons for restraint in using computers as the main alternative to current standardized tests. Haertel (1989), for example, notes that, while "computers are preeminent bookkeepers,. . . existing psychometric models for tailored testing are perilously well suited to the testing of mass of tiny skills." That is, like existing short answer tests., they discourage open-ended situations and ambiguous responses.

An additional problem with the use of computers applies particularly to urban students. Large city school systems serving poor and minority students already have a history of using computers for drill and practice, rather than for creative programming (Ascher, 1984), and these school systems are the most likely to implement rigid uses of computerized systems to meet bureaucratic needs for teacher accountability (Haertel, 1989). Moreover, as Haertel points out,

> teachers in these less affluent systems are often less experienced than those in more attractive positions, reducing the probability that they will successfully adapt or supplement the curriculum to meet their students' needs. Finally, even if they possess the necessary skills to improve upon a test-driven curriculum, teachers in large-city districts may have limited access to supplementary curriculum materials (p. 36).

Thus, unfortunately, it is likely that, with computers, large, bureaucratic urban systems will find it easier to sell old wine in new bottles than to create the real change that is needed so badly.

PERFORMANCE-BASED ASSESSMENT

Research suggests that the best way to gain informative feedback about what students think or are learning in a particular subject discipline is to give them as much range as possible to express themselves fully (Archbald & Newmann, 1988). The best way to measure what students have learned in English class,.for example, may be to ask them to keep a journal of their thoughts after each class, or even to prepare a portfolio of their poems and stories. Similarly, the best way to discover whether third grade students will be able to use what they have learned from a science class may be to probe in a deep and free-ranging way, or even to offer the students a chance to do active science in a laboratory. This kind of active assessment has been called performance assessment, assessment in context, situational testing, or even authentic assessment (Archbald & Newmann, 1988; Brown, et al., 1989; Frederiksen, 1984; Gardner, 1988). At the same time as it maximizes feedback, these kinds of assessment strategies promote instruction geared to the complex thinking that readies students for the messiness of life.

In contrast to standardized tests, where ill-structured problems have always been struck down as unfair, these new assessments intentionally present ill-structured problems. Like life, where most of the important problems one faces are ill-structured, the assessments enable each student to demonstrate mastery in his or her own way. In fact, as in life, students may be given disorderly situations, in which even the problem to solve is not quite clear. Moreover, because the problems and situations are as "ill-structured" as life is, these tests enable students to demonstrate mastery that is meaningful beyond the instructional setting (Archbald & Newmann, 1988).

The most apparent difficulty with such assessments, of course, arises with scoring. Haven't we come full circle to the subjectivity that was once found troublesome in teachers' grades? In fact, NAEP's use of writing assignments suggests that "it is possible to derive reliable, publicly believable quantitative measures from [judgements] of these products" (Resnick & Resnick, 1989, p. 39). Moreover, findings in England and Wales suggest that teachers and external examiners can be trained to score portfolios of students' work with a high degree of agreement (Goldstein & Wolf, 1988). Of course, as American researchers point out, educators here must decide what kinds of performance are valued, and work needs to be done to investigate "just which aspects of student learning to document" (Resnick & Resnick, 1989; Wolf, 1987/1988).

A second objection to performance assessments is that, even if they are scorable, the costs will be exorbitant relative to machine scorable tests. These new assessments are, in fact, more expensive, per pupil. It is impossible to compete with the 10,000 mechanically scored tests per hour ETS now achieves (Frederiksen, 1984). But there is no reason to test as often, or as many students, as is done currently. While all students should be tested when they need to be evaluated or diagnosed, accountability testing does not have to involve every student in every classroom in every state. Sampling methods are now sophisticated enough for states to easily attain quite accurate information on their schools through assessing only a portion of their students. Resnick and Resnick suggest that such sampling methods would keep the cost of mandated testing programs "within tolerable bounds" (1989, p. 47).

DYNAMIC ASSESSMENT

Using the studies of Feuerstein and Vygotsky, both of whom worked with low-achieving children, current researchers are also studying the role of dynamic assessment in instruction. Other ways to refer to dynamic assessment are guided learning, interactive instruction, and reciprocal teaching. Whatever the name, at the most elementary level, the goal of the assessment is not to compare individuals, but to find the instructional approach to help each student reach an acceptable level of performance. Various researchers are working with different procedures: some use "guided learning" to create cognitive maps, and then help students take on greater and greater responsibility for learning. Others offer test/retest reciprocal teaching situations; in this case, the student is either evaluated by how much improvement has occurred with a certain amount of help, or how much help was needed to reach a specific level (Brown, et al., 1989).

Although there is evidence that dynamic assessment techniques improve standardized achievement scores—one study, for example, shows this for bilingual Hispanic students (Duran, 1989)—Brown and her associates warn against reformulating this technique into yet another method of stratifying and sorting. In fact, they point out the educative potential of dynamic assessment:

> To the extent that we can develop truly diagnostic tests of competence in the main academic areas, and to the extent that we can provide more powerful learning environments capable of dealing with those individual differences, it should be possible to avoid the selection and labeling process altogether and instead concentrate on providing help to students when they encounter problems (Brown, et al., 1989, p. 80-81).

CONCLUSIONS

New work in cognition makes clear that both teaching.and testing could be structured to better prepare students for the complex thinking required by life. Since current political trends make it unlikely that the power of testing will decline in our society, or that testing will cease to drive instruction, it is especially important to reformulate assessments so that they can help alter schooling in ways that will effectively and appropriately educate individual students to meet their personal needs as well as those of society.

Because short answer tests have been so important in driving learning in urban schools, and because the size of urban school systems encourages bureaucratic forms of accountability, it will be difficult to create forms of change that demand greater flexibility. However, the new assessment practices offer particular hope to urban students whose gifts and needs are diverse, and who have suffered the most under traditional teaching and testing methods. Portfolios, performance assessments, and the variety of reciprocal teaching methods that rely on dynamic assessment all offer new directions for improving urban education.

REFERENCES

Anastasi, A. (1976). *Psychological testing.* New York: MacMillan.

Archbald, D.A., & Newmann, F.M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school.* Reston, VA: National Association of Secondary School Principals. (ED 301 587)

Ascher, C. (1989). *Urban school finance: The quest for equal educational opportunity.* ERIC/CUE Digest No. 55. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. (ED 311 147)

Ascher, C. (1988a, April). *Summer school, extended school year, and year-round schooling for disadvantaged students.* ERIC/CUE Digest No. 42. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. (ED 298 213)

Ascher, C. (1988b, August). *Grade retention: Making the decision.* ERIC/CUE Digest No. 46. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. (ED 304 498)

Ascher, C. (1984, January). *Microcomputers: Equity and quality in education for urban disadvantaged students.* ERIC/CUE Digest No. 19. New York: ERIC Clearinghouse on Urban Education, Teachers College, Columbia University. (ED 242 801)

Baker, E.L. (1989). Mandated tests: Educational reform or quality indicator? In B.R. Gifford (Ed.), *Test policy and test performance: Education, language and culture.* Boston: Kluwer Academic Publishers.

Bond, L. (1988, December). *Understanding the black/white student gap on measures of quantitative reasoning.* Presented at Testing and the Allocation of Opportunities in Education and Employment to Black Americans, Howard University, Washington, DC. Unpublished manuscript. University of North Carolina.

Bracey, G.W. (1989, May). The $150 million redundancy. *Phi Delta Kappan, 70* (9), 698-702.

Brandt, R. (1988, January). On assessment in the arts: A conversation with Howard Gardner. *Educational Leadership, 45* (4) 30-34.

Brown, A.L., Campione, J.C., Webber, L.S., & McGilly, K. (in press). Interactive learning environments: A new look at assessment and instruction. In B. Gifford & M.C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction.* Boston: Kluwer Academic Publishers.

Camegie Foundation for the Advancement of Teaching. (1988). *Report card on school reform: The teachers speak.* New York: Author.

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the twenty-first century.* New York: Author.

Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C., & Loef, M. (1988, April). *Using knowledge of children's mathematical thinking in classroom teaching: An experimental study.* Paper presented at the annual meeting the AERA, New Orleans, LA. (ED 292 683)

Catterall, J.S. (1987). *Towards researching the connections between tests required for high school graduation and the inclination to drop out of school.* Los Angeles: California University, Center for the Study of Evaluation. (ED 293 886)

Chavers, D., & Locke, P. (1989, April). *The effects of testing on Native Americans.* Prepared for the National Commission on Testing and Public Policy. Unpublished manuscript. Native American Scholarship Fund, Inc., Albuquerque, NM.

Cunningham, A.E. (in press). Eeny, meeny, miny, moe: Testing policy and practice in early childhood. In L.C. Wing & B.R. Gifford (Eds.), *Trends in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Darling-Hammond, L. (1989, Fall). Accountability for professional practice. *Teachers College Record, 91* (1), 59-81.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *The Elementary School Journal, 3*, 133-143.

Dorr-Bremme, D.W., & Herman, J.L. (1986). *Assessing student achievement: A profile of classroom practices.* Los Angeles: UCLA Graduate School of Education, Center for the Study of Evaluation.

Duran, R.P. (1989, October). Assessment and instruction of at-risk Hispanic students. *Exceptional Children, 56* (2), 154-158.

Duran, R.P. (1988, May). *Testing of Hispanic students: Implications for secondary education.* Prepared for the National Commission on Testing and Public Policy. Unpublished manuscript. University of California, Santa Barbara.

Frederiksen, N. (1984, March). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39* (3), 193-202.

Gardner, H. (in press). Assessment in context: The alternative to standardized testing. In B. Gifford & M.C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction.* Boston: Kluwer Academic Publishers.

Glass, C.V., & Ellwein, M.D. (1986, December). Reform by raising test scores. *CRESST Evaluation Comment*, 1-6.

Goldstein, H., & Wolf, A. (in press). Recent trends in assessment: England and Wales. In L.C. Wing & B.R. Gifford (Eds.), *Trends in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Haertel, E. (1989). Student achievement tests as tools of educational policy: Practices and consequences. In B.R. Gifford (Ed.), *Test policy and test performance: Education, language and culture.* Boston: Kluwer Academic Publishers.

Haney, W.M. (1989). Making sense of school testing. In B.R. Gifford (Ed.), *Test policy and test performance: Education, language and culture.* Boston: Kluwer Academic Publishers.

Heyns, B. (1978). *Summer learning and the effects of schooling.* New York: Academic Press.

Jones, L.V. (in press). School achievement trends for black students. In L.C. Wing & B.R. Gifford (Eds.), *Trends in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Jussim, L. (1986). Self-fulfilling prophesies: A theoretical and integrative view. *Psychological Review, 93* (4), 429-445.

Kennedy, M.M., Birman, B.F., & Demaline, R.E. (1986). *The effectiveness of Chapter 1 services.* Washington, DC: U.S. Department of Education, Office of Educational Research and Information. (ED 281 919)

Madaus, G.F. (1989). The Irish study revisited. In B.R. Gifford (Ed.), *Test policy and test performance: Education, language and culture.* Boston: Kluwer Academic Publishers.

Madaus, G., & Haney, W. (in press). *The fractured marketplace for standardized testing.* Boston: Kluwer Academic Publishers.

Massachusetts Advocacy Center. (1990, March). *Locked in/locked out: Tracking and placement practices in Boston public schools.* Boston: Author.

Mehrens, W.A. (1989), Using test scores for decision making. In B.R. Gifford (Ed.), *Test policy and test performance: Education, language and culture.* Boston: Kluwer Academic Publishers.

National Coalition of Advocates for Students. (1985). *Barriers to excellence: Our children at risk.* Boston: Author.

National Alliance of Business (1987). *The fourth R.: Workforce readiness.* Washington, DC: Author. (ED 289 045)

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: U.S. Government Printing Office. (ED 226 006)

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America.* Chestnut Hill, MA: Author.

Neill, D.M., & Medina, N.J. (1989, May). Standardized testing: Harmful to educational health. *Phi Delta Kappan, 70* (9), 688-698.

Oakes, J. (1987). *Improving inner city schools: Current directions in urban district reform.* Santa Monica: RAND Corporation, Center for Policy Research in Education. (ED 291 831)

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven: Yale University Press. (ED 274 749)

OERI State Accountability Group. (1988, September). *Creating responsible and responsive accountability systems.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement. (ED 299 706)

Pennock-Roman, M. (1988, May). The status of research on the scholastic aptitude test (SAT) and Hispanic students in postsecondary education. In L.C. Wing & B.R. Gifford (Eds.), *Trends in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Peters, C.W., Wixson, K.K., Valencia, S., & Pearson, P.D. (in press). Changing statewide reading assessment: A case study of Michigan and Illinois. In L.C. Wing & B.R. Gifford (Eds.), *Trends in educational testing and assessment.* Boston: Kluwer Academic Publishers.

Raizen, S.A., Baron, J.B., Champagne, A.B., Haertel, E., Mullis, I.V.S., & Oakes, J. (1989). *Assessment in elementary school science education.* Andover, MA: National Center for Improving Science Education, The Network, Inc.

Raudenbush, S.W. (1984, February). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76* (1), 85-97.

Reich, R.B. (1988). *Education and the next economy.* Washington, DC: National Education Association, Professional and Organizational Development/Research Division.

Resnick, D.P. (1980). Minimum competency testing historically considered. *Review of Research in Education, 8,* 3-29.

Resnick, L.B., & Resnick, D.P. (1989, March). *Assessing the thinking curriculum: New tools for educational reform.* Prepared for the National Commission on Testing and Public Policy. Unpublished manuscript. Learning Research and Development Center, University of Pittsburgh, & Department of History, Carnegie-Mellon University, Pittsburgh, PA.

Rogers, V. (1989, May). Assessing the curriculum experienced by children. *Phi Delta Kappan, 70* (9), 714-718.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectations and student intellectual development.* New York: Holt, Rinehart & Winston.

Shepard, L. (1989, October). *What is test misuse: Perspectives of a measurement expert. The Users of standardized tests in American Education.* Speech presented at the Invitational Conference of the Educational Testing Service, New York.

Shepard, L. (1988, April). *Should instruction be measurement driven: A debate.* Paper presented at the annual meeting of the AERA, New Orleans, LA.

Slavin, R. (1980). Cooperative learning. *Review of Educational Research, 50,* 315-342.

Sosa, A.S. (1988). *The impact of testing on Hispanics.* The proceedings of a national hearing co-sponsored by the National Commission on Testing and Public Policy and the Intercultural Development Research Association. Berkeley, CA: National Commission on Testing and Public Policy.

Sternberg, R.J. (in press). CAT: A program of comprehensive abilities testing. In B. Gifford & M.C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction.* Boston: Kluwer Academic Publishers.

Teitelbaum, P. (1989). Feminist theory and standardized testing. In A.M. Joggar & S.R. Bordo, *Gender/Body/Knowledge.* New Brunswick: Rutgers University Press.

U.S. Department of Education (1987, July). *White paper on accountability: Tying assessment to action.* Washington, DC: Author.

U.S. Department of Education. (1984). *The nation responds: Recent efforts to improve education.* Washington, DC: Author. (ED 240 748)

Valdes, G., & Figueroa, R. (1989). *The nature of bilingualism and the nature of testing: Towards the development of a coherent research agenda.* Prepared for the National Commission on Testing and Public Policy. Unpublished manuscript. University of California, Berkeley, & University of California, Davis.

Vickery, T.R. (1988, February). Learning from an outcomes-driven school district. *Educational Leadership, 45* (5), 52-56.

Wise, A.E., & Gendler, T. (1989). Rich schools, poor schools: The persistence of unequal education. *The College Board*, 151, 12-37.

Wofford, Sr., D. (1985, December). Report on the implementation of the Basic Skills Assessment Program, 1984-85. Columbia, SC: Department of Education.

Wolf, D.P. (1987 December/1988, January). Opening up assessment. *Educational Leadership, 45* (4), 24-29.

ERIC Clearinghouse on Urban Education

ERIC Clearinghouse on Urban Education
Institute for Urban and Minority Education
Box 40, Teachers College, Columbia University
New York, New York 10027