

FOCUS ON EXCEPTIONAL children

Performance Assessment and Special Education: Practices and Prospects

James A. Poteet, Joyce S. Choate, and Susan C. Stewart

The field of educational assessment is changing rapidly and dramatically. At the center of change is performance assessment. The questions that frame this issue are:

- What is performance assessment?
- How is it currently used?
- What is its value to special education?
- What is its future?
- What should special educators do to best accommodate students with disabilities?

FORCES INFLUENCING ASSESSMENT CHANGES

After decades of widespread traditional standardized testing in the schools, this practice is finally changing. Several forces have converged to both create the need and provide the direction for change: educational reform, national standards, changing curricula, and outright rebellion against traditional testing.

Educational Reform and Assessment

Educational assessment was jolted out of a state of complacency by the success of Sputnik in 1957. Since then, testing has been used both to justify and to decry the status of education in America's schools. Following a decade of emphasis on science and mathematics so America's students would be prepared to meet the challenges of national security, the National Assessment of Educational Progress (NAEP) test was developed in 1968 to serve as a report card for the entire nation. In spite of the push for higher achievement in critical subject areas, the 1970s saw a decline in scores on SAT and ACT college admissions tests. The public's reaction and call for school curricula to "return to the basics" was not unexpected, nor were new tests to measure minimum competencies in at least 39 states.

Following publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983), America issued the call to restructure education for excellence. Many states raised educational standards and began to develop statewide testing programs, expanding existing minimum competency tests to include more grade levels and subject areas, and to use the test results as part of high school graduation requirements.

James Poteet is a professor of special education at Ball State University. *Joyce Choate* is a professor of curriculum and instruction at Northeast Louisiana University. *Susan Stewart* is an associate professor of education at the University of West Florida

The 1990s cast educational testing as the central focus for the educational reform preached by political leaders and policy-makers. Movement toward a global economy with its international competition as the Millennium approaches has sparked political, economic, and educational hysteria concerning the knowledge level of America's students.

National Educational Standards and Assessment

Reform movements reached a fervor with the Bush administration's *America 2000* (U.S. Department of Education, 1991), a project resulting from an "Education Summit" meeting of all governors in 1989. "Specific results-oriented goals" and "accountability for outcome-related results" were mentioned as popular reforms to pave the way for a nationally established system of educational standards and assessment. The National Education Goals Panel, created in 1990 to report annual progress toward the goals, recommended creating national education standards and identifying a system to assess the standards. Following this recommendation, Congress established the National Council on

Education Standards and Testing (1992), charged with providing "advice on the desirability and feasibility of national standards and testing in education" (p. B-1). The NCEST report, *Raising Standards for American Education*, endorsed "the adoption of high national standards and the development of a system of assessments to measure progress toward those standards" (p. 8). This report was met with an opposing statement by members of the Coalition of Essential Schools, the National Education Association, American Educational Research Association, and private individuals urging caution in the development of a national testing system. Koretz, Madaus, Haertel, and Beaton (1992) offered congressional testimony against some of the recommendations made by the National Council on Education Standards and Testing, pointing out many of the difficulties in adopting alternative approaches to assessment.

Ravitch (1993), who held a leadership role in developing the national standards and assessment in the Department of Education in 1991 and 1992, noted that the new assessment program was to be developed by states and to be based on national standards. The federal role would be to provide funds for states and appropriate organizations to develop curricular standards, with input from people most familiar with subject matter content. The work of the National Council of Teachers of Mathematics served as a model for other organizations. Grants were awarded to a number of professional organizations to set standards in various disciplines, with the final draft of national standards to be completed by 1994-1995.

Possible Uses of a National Assessment System

A national assessment system "eventually could be used for such high-stakes purposes for students as high school graduation, college admission, continuing education, or certificate for employment. Assessments could also be used by states and localities as the basis for system accountability" (National Council on Education Standards and Testing, 1992, pp. 27-28). The NCEST council further noted its interest in performance-based assessments, portfolios, and projects as means of collecting information about what students know and can do. The national assessment, then, obviously is a system of high-stakes testing for accountability purposes. The interest in large-scale national assessment does not stop with the boundaries of the United States, as evidenced by topics such as performance and large-scale assessment featured at the International Test Commission conference in the summer of 1993 (Hambleton & Oakland, 1993).

One approach toward a national assessment system would be to draw upon the assessment program currently in use, the National Assessment of Educational Progress (NAEP) tests. NAEP's agenda will address the ideas of reg-

FOCUS ON EXCEPTIONAL CHILDREN

ISSN 0015-511X

FOCUS ON EXCEPTIONAL CHILDREN (USPS 203-360) is published monthly except June, July, and August as a service to teachers, special educators, curriculum specialists, administrators, and those concerned with the special education of exceptional children. This publication is annotated and indexed by the ERIC Clearinghouse on Handicapped and Gifted Children for publication in the monthly *Current Index to Journals in Education* (CIJE) and the quarterly index, *Exceptional Children Education Resources* (ECER). It is also available in microfilm from Xerox University Microfilms, Ann Arbor, MI. Subscription rates: Individual, \$30 per year; institutions, \$40 per year. Copyright © 1993, Love Publishing Company. All rights reserved. Reproduction in whole or part without written permission is prohibited. Printed in the United States of America. Second class postage is paid at Denver, Colorado. **POSTMASTER:** Send address changes to:

Love Publishing Company
Executive and Editorial Office
1777 South Bellaire Street
Denver, Colorado 80222
Telephone (303) 757-2579

Edward L. Meyen
University of Kansas

Glenn A. Vergason
Georgia State University

Richard J. Whelan
University of Kansas Medical Center

Stanley F. Love
Publisher

Holly T. Rumpler
Senior Editor

ularly collecting state-level data, setting achievement levels, rating student performance, and linking local district, state, and national results with international test results. It will urge Congress to let school districts use NAEP information at their own discretion and costs ("NAGB Adopts Policy," 1993). The National Education Goals Panel suggested that the NAEP expand to additional grade levels and subjects and use results to compare schools and monitor achievement toward national standards and goals. Thus, NAEP has a key role in the national assessment movement.

Special Education Students and a National Assessment System

The recommended standards and assessments seem to apply to *special education* students. According to Ravitch (1993), "the [standard-setting] projects are *not* to create standards for the elite, but to create them for *all* children. The projects are *not* to create a national curriculum, but to describe what children should know and be able to do in a particular field" (p. 771). The NCEST council has stated that both the standards and the assessments could boost high- and low-achieving students' performance. "Students with disabilities or of limited English proficiency should be provided opportunities to learn and to demonstrate their mastery of material under circumstances that take into account their special needs" (National Council on Education Standards and Testing, 1992, p. 10).

This seems to be in stark contrast to current practice. In NAEP assessment, from 2 to 8% of the total sample is excluded because of disability or limited English proficiency. A report from the National Center on Educational Outcomes by James Ysseldyke and Kevin McGrew concluded that "about 40 percent to 50 percent of students with disabilities are excluded from national assessments" (in Moses-Zirkes, 1992, p. 45).

As plans develop for a system of national assessment, the inclusion of special education students is under consideration for the future. The U.S. Department of Education, Office of Special Education Programs (OSEP) is collaborating with a task force of professionals to increase inclusion of exceptional students in the testing. At the same time, it is aware that alternative assessments may be difficult for some students because of complex language and abstract content. For example, some tasks could be inappropriate for a student with dyslexia who cannot read written instructions or a student with hearing impairment who cannot hear oral directions ("Record Number of Students," 1993). Clearly, as movement toward national standards and assessment gains momentum, decisions about including special education students in national, state, and local assessment practices are as important as decisions about including them in instruction in general education programs.

Changing Curricula

Efforts to improve our current educational system have shifted to initiatives aimed at the fundamental redesign of schools and of approaches for teaching and learning. Among the initiatives reflected in changing curricula are the whole language emphasis, the interdisciplinary curriculum, team and collaborative teaching, cooperative learning, adaptations for inclusive practices, an emphasis on authentic instruction and real-life tasks, a curriculum centered on developing higher-order performances and cognitive skills, and the use of alternative assessments.

Darling-Hammond (1993) emphasized the importance of ensuring that *all* students learn at high levels. She discussed the need for teachers to adapt and respond on the basis of individual needs and interactions to complex, ever changing circumstances. Therefore, effective teaching techniques should vary for different students, for different subject areas, and for different instructional goals.

Curricular changes are impacting assessment practices, and the reverse also is true. Coutinho and Malouf (1993) noted that the increasing use of alternative performance assessment is expected to redirect curriculum and instruction toward current and more holistic theories of learning. As Wiggins (1989) aptly described it: "If tests determine what teachers actually teach and what students will study for—and they do—then the road to reform is a straight but steep one: test those capacities and habits we think are essential, and test them in context" (p. 41).

Rebellion Against Traditional Testing

Along with the demand of accountability by decision-makers came *rebellion* against the use of tests to measure that accountability. Critics of standardized testing have charged that the tests do not reflect accurate information about learning; are biased against minorities, people with limited proficiency in English, and low-income families; influence teachers to use class time to prepare students to take the tests; and focus instruction on segmented skills instead of higher-order thinking and creativity (Haney & Madaus, 1989). As professionals noted their resistance against using multiple-choice tests as the sole method to determine achievement, some states began trying new forms of tests to get a more realistic, authentic picture of student knowledge (Educational Testing Service, 1990).

Problems with Traditional Testing

The National Commission on Testing and Public Policy (1990), funded by the Ford Foundation, issued a report, *From Gatekeeper to Gateway: Transforming Testing in America*, noting that the testing movement must be trans-

formed so it develops human potential and allocates opportunities from kindergarten through the workplace. This report and other sources (e.g., Haney & Madaus, 1989; Livingston, Castle, & Nations, 1989) and practical experience point out the following major problems with standardized testing:

1. *Misleading information.* The results from traditional tests provide a limited view of student learning. Choate & Evans (1992) noted that many classroom skills are not assessed by traditional tests because the tests sample a broad range of skills. Test results can mislead the consumer because many people think the score is an absolute rather than merely an indicator. Furthermore, test scores do not explain the approach students take when responding to test items. Some students who can perform a given test item successfully fail the item for reasons not related to their knowledge of the correct answer (e.g., test anxiety, feeling sick or sleepy, effects of medications, or difficulty marking answer sheets). Thus, using these tests to make decisions about students' futures (e.g., if they will pass reading) may result in incorrect decisions. A "good" (valid) test has good (accurate) predictive validity; many standardized norm-referenced tests do not.
2. *Unfairness.* Many of the traditional tests are unfair to certain populations in America, particularly some students with disabilities (Choate, Enright, Miller, Po-teet, & Rakes, 1992). They are biased in the sense that the language usage, cultural examples, and learning style or skills required to successfully complete test items do not reflect those of populations such as people with limited English proficiency, low incomes, and membership in certain minority or cultural groups and females. Some tests tend to favor middle- to upper-class white males (Neil & Medina, 1989).
3. *Distraction.* Test results often are used as measures of accountability—for example, in reaching reading objectives and goals and as yardsticks for comparing schools to determine which is "better" (has higher average scores). When these factors are emphasized, attention is distracted from social issues that may contribute to low average scores—overcrowded classrooms, lack of adequate instructional materials, administrative focus on test scores rather than on acquiring knowledge, high-level thinking, and so on.
4. *Quality control.* Testing is big business in America. The taxpayer cost of state and local testing is approaching \$1 billion a year. Students at some grade levels are taking as many as 12 comprehensive tests in a single year (Moses, 1990). However, there is no

authority to guarantee regulations or quality control, although recent demands for alternative assessment may eventually result in quality control.

5. *Expense.* Traditional testing is quite expensive in terms of cost of materials and time (administering, scoring, posting results). The money allocated for traditional testing in a school budget might best be allocated to other approaches to assessment that are less expensive, more time-efficient, less biased, and produce results that have more meaning for educational decisions.
6. *Use of results.* Standardized tests all too often are used unfairly to shut Americans out of education and employment opportunities (Moses, 1990). In an interview with Kirst (1991), Lorrie Shepard pointed out that when tests drive instruction, the curriculum often covers only what is on the test. Teaching to the test—at least when the test is a standardized multiple-choice test—restricts the curriculum to covering test items that are rather factual, concrete, and isolated from a broader integrated scope of desired curricula. Hanson (1993) portrayed tests as the gatekeepers, allowing only some individuals to pass through. He suggested that test scores can redefine an individual in the eyes of the individual and in the eyes of others.
7. *Sending the wrong message to students.* Traditional tests foster a one-right-answer mentality, limiting students' thinking processes. Although some knowledge actually has only one correct answer, other knowledge (e.g., causes of the Civil War), has no one right answer (Hambleton & Maurphy, 1991).

The Cry for a Different Kind of Test

Reactions to the problems associated with traditional testing resulted in a search for more appealing alternatives. Reporting on congressional hearings on March 7, 1991, Lauren Resnick criticized traditional assessment because of its focus on skills and ability rather than on the curriculum. She recommended an overhaul to include performance examinations, portfolios, and curriculum-related projects tied to real-life demands (American Psychological Association, 1991). Almost as soon as President Bush made public the *America 2000* plan, criticisms of it were published, especially regarding the assessment plans. In citing school districts (e.g., Pittsburgh and Albuquerque) and states, (e.g., California, Connecticut, and Maryland) that have been developing and implementing newer forms of assessment, Darling-Hammond (1991) put it well: "Recognizing these problems [with tests], many schools, districts, and states have recently begun to develop different forms of assess-

ment [that] require students to think analytically and demonstrate their proficiency as they would in real-life performance situations” (p. 16).

Other professionals, such as curriculum specialists, policy-makers, and teachers, advocate replacing multiple-choice tests with assessments of oral reports, exhibitions, projects, portfolios, performance assessments, writing samples, observations, self- and peer-assessment, reviews, and the like. In essence, these advocates want to see assessment more closely aligned with classroom instruction (Hambleton & Murphy, 1991).

Congress asked its Office of Technology Assessment (1992) to prepare a report about testing in this country. The report, *Testing in American Schools: Asking the Right Questions*, “documents the central role of educational testing in recent national debates about educational reform” (Linn, 1993, p. 1). With the strong movement toward a national assessment program, many professionals began to wonder about the best type of assessment approach to take and what the results of such a movement might be.

The public wanted more than simple test scores as evidence of what students could actually do. Minority groups were concerned about discrimination in test items. Ravitch (1993) noted that “unless current tests change, the [national] standards will wither and die. Teachers know that they will be judged by test results, and they will continue to teach to the tests by which they are judged” (p. 772).

Given the dissatisfaction with traditional standardized multiple-choice tests and the desire for alternative approaches to assess student knowledge for purposes of accountability, improved instruction, and program planning, a wave of change is beginning to sweep the assessment field as alternative approaches to assessment take shape and gradually are implemented in the classroom.

PERFORMANCE-BASED ASSESSMENT: THE ALTERNATIVE TO TRADITIONAL ASSESSMENT

Alternatives to traditional standardized assessments are necessary to help teachers answer the questions that will help solve their instructional needs: What has the student learned? How has knowledge been acquired? How is knowledge used? What levels of knowledge—low, concrete, factual level or higher-order thinking, problem-solving level—have been learned? Traditional tests do not give teachers that information. Teachers are demanding change in assessment practices to give them information on which to base instructional decisions.

Politicians and other policy decision-makers are looking for ways to improve the ranking of America’s students in

national and global comparison. Evolving out of these concerns is a new name to describe assessment that is radically different from traditional, standardized, norm-referenced tests: *alternative assessment*. With the birth of this new terminology come the throes of definition and characteristics.

Alternative Assessment

Alternatives to standardized testing have been described by a number of terms, including *alternative assessment*, a somewhat generic title, *direct assessment*, *performance assessment*, *authentic assessment*, and *portfolio assessment*. Worthen (1993) contends that, despite the subtle differences in emphasis indicated by the descriptors, “all exhibit two central features: first, all are viewed as *alternatives* to traditional multiple-choice, standardized achievement tests; second, all refer to *direct* examination of student *performance* on significant tasks that are relevant to life outside of school” (p. 445). Several methods of applied behavior analysis, such as observation and evaluation of permanent products, may be incorporated (Elliott, 1992). Descriptions also may emphasize the criterion-referenced nature of alternative assessment and note defining features such as authentic tasks that “represent actual progress toward instructional goals” (Pierce & O’Malley, 1992, p. 1).

A major difference between traditional and alternative assessment is the type of response required by the student. Standardized, norm-referenced group tests require a student to *select* and mark correct responses. Alternative assessment approaches, on the other hand, require the student to *produce*, construct, demonstrate, or perform a response. The demonstration of knowledge often is required to be in a format that allows others to understand the process or the product, or both.

Thus, alternative assessment is a generic term encompassing several categories of assessment. Three types of alternative assessment are considered here: performance-based (or simply performance), authentic, and portfolio assessment.

Performance Assessment

Performance tests are designed to assess what the student can *do* with knowledge, rather than the isolated specific *bits* of knowledge the student possesses. A performance test should be designed so that simply memorizing facts or concepts is not requested, as often is the case in simple open-ended or fill-in-the-blank test items. To be a performance test, it must require that the student *do* (produce, demonstrate, perform, create, construct, apply, build, solve, plan, show, illustrate, convince, persuade, or explain) some task. In some situations, successful performance of the task is required for promotion to the next grade, graduation, employ-

ment, admission to college, or other "high-stakes" decision. The task requirement in a performance test is quite different from that typically found in a traditional, standardized, multiple-choice test on which the student is to know or guess the correct answer and *select* it.

As indicated by the sample descriptions of a performance task below, although exact definitions may vary slightly, they all reflect the notion of demonstrated knowledge and skill. *Performance* consistently refers to the type of response the assessment task requires (Meyer, 1992). A performance task:

- requires direct demonstration of skills (Jacobs & Chase, 1992).
- involves doing things rather than knowing things (Carroll & Hall, 1985).
- "reflects student performance on instructional tasks" (Pierce & O'Malley, 1992, p. 2).
- requires a student to demonstrate knowledge or skills by constructing a product or answer (Office of Technology Assessment, 1992).
- allows students to perform a complex skill or procedure or create a product to demonstrate that they can apply the knowledge and skills they have learned (Jacobs & Chase, 1992).
- "is an exercise in which a student demonstrates specific skills and competencies in relation to a continuum of agreed upon standards of proficiency or excellence" (Pierce & O'Malley, 1992, p. 2).
- "covers many different types of testing methods that require students to demonstrate their competencies or knowledge by creating an answer or a product" (Feuer & Fulton, 1993, p. 478).
- "simultaneously requires the use of knowledge, skills, and values that are recognized as important in a domain of study" (Gitomer, 1993, p. 244).
- "is best understood as a continuum of formats that range from the simplest student-constructed responses to comprehensive demonstrations or collections of large bodies of work over time" (Feuer & Fulton, 1993, p. 478).
- "is qualitatively consistent with tasks that members of discipline-based communities might conceivably engage in" (Gitomer, 1993, p. 244).
- "depends heavily on professional judgment [for evaluation of performance]" (Linn, 1993, p. 9)
- "entails judgments and reports of the quality of performance by community members" (Gitomer, 1993, p. 244).
- "relies on professional rater judgment in its design and interpretation" (Pierce & O'Malley, 1992, p. 2).

- allows the evaluators to observe the finished product as well as the procedure used to solve a problem or create a product (Jacobs & Chase, 1992).

Performance tests are not new. Vocational education has used them for years. Siegel (1986) noted that Adkins described the *work sample method* of performance testing in 1947, and since then little has changed. "Various types of performance assessments were the norm before the introduction of multiple-choice testing in this country and remain the norm in many other countries. Performance assessment is also a part of day-to-day classroom activities for many teachers" (Linn, 1993, p. 9). What is new is the attempt to systematize and apply performance assessment, particularly authentic performance assessment, on a grand scale.

Authentic (Performance) Assessment

Performance assessment becomes *authentic* when it requires realistic demands and is set in real-life contexts (Meyer, 1992). Although the real-life context for assessment tasks ultimately must focus on nonschool settings, the classroom also is a real-world setting, particularly for younger students. Explanations of the relationship between performance and authentic assessment are not always consistent (Coutinho & Malouf, 1993) and the terms are sometimes used interchangeably. Even so, the view reflected throughout the discussions here is that authentic assessment is the "realistic" subset of performance assessment.

Elliott (1992) characterized authentic assessment as a "neo-behavioral approach to classroom assessment" and portrayed it as both a philosophy and a methodology of classroom assessment. Coutinho and Malouf (1993) stipulated that tasks must be real examples of "valued performances."

Authentic assessment can be considered a *concept* about assessment rather than a particular *type* of assessment. Essentially, any type of assessment can be authentic if it focuses on *application* of knowledge to real-life, real-world settings. To ask a student to play a piano piece is more authentic than to ask the student to fill in bubbles on an answer sheet about notation, rhythm, and pitch, even though this knowledge must be applied for successful performance. According to Archbald and Newmann (1988), valid assessment tasks are authentic when they are worthwhile, significant, and meaningful.

Authentic assessment originated in the arts, in which actual performance provides the opportunity for assessment. For example, to evaluate a musician's abilities and progress, we listen to the student perform and analyze what we hear and see. Then, for instructional purposes, we offer feedback, coach, monitor, adjust instruction, and assess performance. Authentic assessment also is appropriate for evalu-

ating actual performance in all curricular areas (National Center for Fair and Open Testing, 1992).

Wiggins (1989) offered the following description of the four basic characteristics of authentic tests:

1. They are designed primarily to represent performance in real-world settings. Once the task has been decided, the logistics of scoring, reliability, and other technical matters can be considered.
2. The criteria for attaining stated goals or objectives are given more attention than usual during the teaching and learning process.
3. The skills of student self-assessment play a greater role than usual.
4. To assure mastery, students should be required to present their work in a variety of settings and formats, publicly and orally, to defend the quality and the content.

Authentic measurement seems to aim toward making assessment more closely resemble actual learning tasks and permitting assessment of higher-order cognitive skills (e.g., problem-solving, critical thinking, reasoning). Of special importance to many advocates of authentic measurement is its capacity for permitting multiple correct answers or multiple acceptable methods to solve a problem or to complete a task.

Authentic performance tests have been and continue to be the assessment method of choice in vocational education (Siegel, 1986). Three types of performance assessment in vocational education, as described by Slater (1980), provide examples of authentic performance assessment procedures: (1) direct assessments, (2) work samples, and (3) simula-

tions. These three formats apply equally to a variety of students and settings.

1. *Direct assessment.* The examinee is evaluated in the actual target context in response to the environment and the evolving situation. "Situation assessment" requires systematically observing a client in an actual work situation (Sitlington, 1979).
2. *Work sample.* A representative task from some work situation is established. Mastery of the task is interpreted as mastery of the skills needed for actual job performance. Scoring is based on performance quality (Siegel, 1986).
3. *Simulation.* A task and the setting are contrived to represent the appearance, form, and effect of a real-life (authentic) situation. Simulations have been particularly valuable in obtaining language samples in classrooms (Shohamy, 1992).

An array of examples of authentic assessment is presented in Table 1. As can be seen in these examples, essential tasks are assessed in real-life contexts. Authentic performances also may be defended and self-evaluated. The final generic category includes examples of authentic assessment spanning all content areas.

Portfolio Assessment

Portfolios typically contain the observable *evidence* or products of performance assessment—evidence that may or may not reflect authentic tasks. When a portfolio includes demonstrations of authentic tasks, that portfolio represents a type of authentic performance assessment. In addition, according to Worthen (1993), portfolios are valuable assess-

TABLE 1
Examples of Authentic Assessment

<p>Reading</p> <ul style="list-style-type: none"> • Actual or audio/video tape of reading to peer • Log and critiques of books read • Book review • Book jacket design 	<p>Oral Expression</p> <ul style="list-style-type: none"> • Transmission of message to several classes • Phone call to request information • Debate on current issues • Persuasive speech 	<p>Written Expression</p> <ul style="list-style-type: none"> • Student interviews • Article for school paper • Written resumé and job application • Invitation to party • Letter to editor 	<p>Mathematics</p> <ul style="list-style-type: none"> • Solving real-life problems using math knowledge • Solving a puzzle using logic and reasoning • Monitoring a savings account • Personal budget
<p>Science</p> <ul style="list-style-type: none"> • Scientific experiment to prove theory • Original investigation and report of findings • Journal of observations of moon, stars • Solutions to local environmental problems 	<p>Social Studies</p> <ul style="list-style-type: none"> • Map of school or community • Design of museum exhibit on topic of interest • Advertising campaign for political candidate • Identification of social problem for co-op group 	<p>The Arts</p> <ul style="list-style-type: none"> • Design and decoration of bulletin board • Submission of art to contest • Artwork design for public building • Performance in a play 	<p>Generic</p> <ul style="list-style-type: none"> • Reflective journal of learning progress • Competition for "grant" money • Planning and teaching a lesson to peer • Laser disk storage of assessment information

ment *tools* because they provide opportunities to both record and review performance data.

Portfolio assessment "is the use of records of a student's work over time and in a variety of modes to show the depth, breadth, and development of the student's abilities; [it] is the purposeful and systematic collection of student work that reflects accomplishment relative to specific instructional goals." (Pierce & O'Malley, 1992, p. 2). Included in the portfolio may be a student's responses to performance as well as traditional assessments, and the student's self-evaluation. Portfolios recently have become especially popular in the area of written expression. Included is a variety of genres and successive drafts so progress can be evaluated. "The great strength of a portfolio is that it obliges us to face squarely the very difficult question of what constitutes high-quality work" (Abruscato, 1993, p. 475).

Portfolios are not new to all areas. They have been standard means of evaluating works of photography, musical compositions, art, dress designs, and the like for centuries. Today, videotapes provide a new approach to collecting examples of one's work.

The purpose of portfolios and how they are to be used are important points to consider before implementing them. Nolet (1992) differentiated *assessment* and *instructional* purposes of portfolios. The "assessment portfolio" contains systematically collected information to help the teacher make educational decisions to verify the effectiveness of instruction, whereas the "instructional portfolio" is used for motivation, facilitating discussions, or promoting reflection to help the student perform new behaviors. Portfolios have the advantage of providing an alternative assessment for various types of assignments, for evaluating a set of revision skills, for encouraging collaboration between student and teacher, and for allowing the teacher to intervene when necessary to provide new direction when the need for such direction becomes apparent in midstream (Jacobs & Chase, 1992). *Assessment* portfolios contain samples of student work over time and under varying conditions relating to a central content area. The work is relative to authentic conditions and incorporates the teacher's commentary or evaluations. Contents of the *instruction* portfolio are decided by both the student and the teacher (Nolet, 1992).

The purpose of the portfolio should determine its contents. If it has no purpose, the portfolio is simply a folder of student work. Portfolios should make sense of students' work, with the contents in chronological order, dated, and organized by category. Grady (1992) suggested maintaining two types of portfolios: *working* portfolios, which are in progress, and *best* or *permanent* portfolios. Evaluation should compare current work with earlier work to indicate progress toward developmentally appropriate standards set

forth in the curriculum. Portfolios are not meant to be used to compare children but, instead, to inspect progress over time. They are good to show parents at conference time (Grace, 1992).

The writing assessment portfolio in Vermont (grades 4 and 8) contains a required writing assignment and six various types of pieces written during the academic year, (e.g., poem, letter, play, and the like) (Abruscato, 1993). The Vermont mathematics portfolio reflects achievement in problem-solving and communication. Contents might include a puzzle, an investigation, an application, and other pieces of mathematics works. "Portfolios also might contain required information for state- or district-wide systems, but these data need not dominate or divert portfolio assessment from being used to inform classroom instruction" (Pierce & O'Malley, 1992, p. 2).

Other examples of portfolio contents are presented in Table 2. As can be seen, the exact content of a portfolio depends upon the subject or topic for which it is used. In general, a portfolio contains observable evidence of a student's progress—a range of selected work samples that demonstrates growth over time.

No set rules exist on how to evaluate the contents of a portfolio. Linn and Baker (1992b) expressed concerns about scoring portfolios and establishing common standards when they are used for accountability. Most important, they asked if portfolios should be scored at all, considering the associated concerns of reliability, validity of rubrics, cost, and time. In a preliminary study in Michigan, they found that descriptive review of portfolios can serve accountability purposes at reduced cost. This approach also would seem to maintain the initial purpose and use of the original portfolio—as a showcase for an individual's best and representative works.

Portfolios have been termed "time-intensive," meaning they require a large amount of teacher time and effort. Still, many teachers, especially in Vermont and Michigan, believe the portfolio has improved their instruction (Dietel, 1992).

Performance Assessment in Perspective

More than 20 years ago, performance contracting was popular in education. During that time, the Multi-State Consortium on Performance-Based Teacher Education was publishing its newsletter concerning issues in teacher training. Esin Kaya (1974) wrote a short article about the competency-based teacher training program at Hofstra University. In discussing how to evaluate teachers, the writer cited behavior demonstrated in a natural setting as the best evidence of mastery of competencies; performance in a simulated setting was listed as a second source of assessment data. Students' products—lesson plans, projects, curricular materi-

als—were described as a third source of evaluative information. These “demonstrations of competency” identified 20 years ago describe the current definitions of performance, authentic, and portfolio assessment, respectively!

CURRENT USE OF PERFORMANCE ASSESSMENT

Performance assessment is gaining ground rapidly in general education, whereas its application in special education is much more restrained. One must remember, though, that when performance assessment is discussed, it may be in relationship to authentic assessment or portfolio assessment, or it may be termed alternative assessment, or it may be used to mean a combination of some or all of these.

Performance Assessment in General Education

Performance assessment is used in several areas of general education. A number of statewide assessment programs feature performance assessment as a method of evaluating the achievements of general education students. Performance assessment is used routinely for evaluating certain competencies. Publishers of commercial tests are beginning to incorporate performance features into their products, and special projects currently are exploring the use of performance assessment.

State Requirements

Various types of performance assessment are being used throughout the states. Some are statewide, and others are confined to a single school system within a state. The Office of Technology Assessment (1992) reported that more than half the states are using some form of performance assessment. The exact nature and extent of implementation varies though.

Assessment of Competency

Performance tests may be used to assess fluency in oral expression through oral interviews, story retellings, directed dialogues, incomplete stories that students are to finish, teacher observation checklists, and student self-evaluations. For large-class assessment, students may be evaluated in pairs or in small groups (Pierce & O'Malley, 1992). Assessment for credentials in certain subject areas, such as foreign languages, may include performance tasks (Wesche, 1992).

Commercial Tests

Commercial test publishers have incorporated the concepts of performance assessment into some series of tests. For example, the Psychological Corporation calls its alternative assessment group the *Integrated Assessment System*. Problem-solving strategies, logical reasoning, and concept

TABLE 2
Examples of Assessment Portfolios

Reading Portfolio

- Audiotape of oral reading of selected passages
- Original story grammar map
- Transcript of story retelling
- Log of books read with personal reactions, summaries, vocabulary
- Representative assignments; responses to pre/post-reading questions
- Favorite performance
- Journal entries including self-evaluation

Science Portfolio

- Representative work samples
- Student-selected best performance
- Report from hands-on investigation
- Notes on Science Fair project
- Journal entries including self-evaluation

- Learning progress record
- Report cards
- Personal journal

Writing Portfolio

- Scrapbook of representative writing samples
- Selected prewriting activities
- Illustrations/diagrams for one piece
- Log/journal of writing ideas, vocabulary, semantic maps, compositions, evaluations
- Conference notes, observation narratives
- Student-selected best performance
- Self-evaluation checklists and teacher checklists

Social Studies Portfolio

- Representative work samples
- Student-selected best performance
- Design of travel brochure, packet or itinerary of trip
- Notes on History Fair project
- Journal entries including self-evaluation

Generic Portfolio

- Tests
- Significant daily assignments
- Anecdotal observations

Mathematics Portfolio

- Reports of mathematical investigations
- Representative assignments
- Teacher conference notes
- Descriptions and diagrams of problem-solving processes
- Video, audio, or computer-generated examples of work
- Best performance
- Journal entries including self-evaluation

Arts Portfolio

- Best performance
- Favorite performance
- First, middle, and final renderings of projects
- Tape of performance
- Journal entries including self-evaluation

- Photographs
- Awards
- Personal goals

development are emphasized in the performance assessment tasks.

The *Language Arts Assessment Portfolio*, published by American Guidance Service (AGS) in 1992, is billed as a "system to document student growth." It includes both portfolio and performance assessment measures at three levels. Self-assessment also is an important component.

Special Projects

Various research reports dealing with alternative and performance assessment are available for a slight charge from the Center for Research on Evaluation, Standards, and Student Testing, UCLA Graduate School of Education. This is an excellent source of information about special projects that are currently active and those that have been completed recently.

Performance Assessment in Special Education

Because of limited implementation, discussions of performance assessment in special education tend to be somewhat speculative. Similarly, the role of special educators in its development or implementation is unclear. Study of these very issues defines the purpose of a working group, composed of officials in the Department of Innovation and Development (DID), Office of Special Education Programs (OSEP), and the ERIC/OSEP Special Project (Coutinho & Malouf, 1993).

According to a recent survey of state assessment practices by Roeber, Bond, and van der Ploeg (1993), even though most states include one or more performance components in their statewide assessment programs, their provisions for students with disabilities vary. For example, in California, Georgia, Kentucky, and Vermont, students with disabilities are included routinely in statewide performance assessments; in Florida these students are regularly excluded. A number of states, including Arkansas, Arizona, Delaware, Louisiana, and Virginia, rely on each student's IEP to determine inclusion or exclusion from state assessments. State policies provide for a number of different situations, as outlined in Table 3. The demands that state performance assessments place upon special students to achieve also vary according to state or school reporting and interpretation policy. For example, although Maryland's students with disabilities may be included in state performance assessments, they do not *have to* master the tasks because the scores of schools rather than individual students are identified.

As noted previously, performance assessment has long been an integral part of vocational assessment. Variations of performance assessment and instruction (e.g., coached simulations and approximations) have been common practices in many special education programs, especially for students

preparing for transition to a less restrictive instructional setting, a given job, or independent living. In addition, according to Coutinho and Malouf (1993), the *principles* underlying performance assessment historically have prevailed in popular special education assessment practices such as behavioral, curriculum-based, direct, ecological, functional, and naturalistic assessment.

CRITIQUE OF PERFORMANCE ASSESSMENT

Proponents of performance assessment have touted a number of general advantages of performance assessment. As would be expected, critics cite many disadvantages. When considered for use with students with special needs, however, identifying positive and negative aspects of performance assessment is confounded somewhat by limited implementation.

General Advantages

Some believe that performance tests avoid the deleterious effects of accountability associated with norm-referenced tests. The advantages of performance assessment may be seen as opposites of the criticisms of traditional multiple-

TABLE 3
Examples of Policy Variations
for Including Students with Disabilities
in Statewide Performance Assessments

Kansas	Excluded unless IEP calls for inclusion
Delaware, Ohio, New Jersey, Tennessee	Included unless specifically excluded by IEP
Texas	Special committees decide if individual students are included
Idaho, Hawaii, Rhode Island	Students mainstreamed for at least 50% of school day are included
Pennsylvania	Students mainstreamed for reading instruction are included
New York	Alternative testing procedures are permitted
New Mexico	Modifications of test procedures and waivers are permitted
Massachusetts	Excluded at parents' request or by IEP

Information extracted from E. Roeber, L. Bond, and A. van der Ploeg (1993). *State Student Assessment Program Data Base*. (Washington, DC: Council of Chief State School Officers/North Central Regional Education Laboratory.)

choice tests. The former focus on curricula instead of skills and abilities; the tasks are related to real-life demands and require students to really think, not memorize; and the results are instructionally useful and can be used for program planning.

A number of positive features are cited throughout the literature. For example, Kamphaus (1992) categorized performance assessment as an approach aimed at improving educational assessment practices. Coutinho and Malouf (1993) note its congruence with current learning theories, while Archbald (1992) cites congruence of school-based authentic assessment with curricula. Christenson (1992) found it consistent with research-based effective instruction.

Performance assessment also is useful for monitoring performance to improve instruction. Among the possible uses of portfolio and other performance assessment results are making diagnosis and placement decisions, monitoring student progress, evaluating instructional effectiveness, providing feedback to students, and communicating with other teachers, specialists, and parents (Pierce & O'Malley, 1992).

Potential for Special Students

The specific benefits of performance assessment for students with disabilities are yet to be fully realized. Many of the general advantages cited also apply to special learners as well. In addition, both preliminary reports of use in general education and the nature and characteristics of performance assessment suggest several possibilities. Choate and Evans (1992) pointed out that authentic assessment practices might offer students with disabilities (a) multiple options for demonstrating knowledge, (b) more instructional time if assessment is integrated with instruction, (c) an accurate measure of curriculum attainment, (d) knowledge of relevant skill acquisition, as both instruction and assessment would be related directly to essential survival skills, (e) self-monitoring strategies and student habits if self-assessment approaches are built in, and (f) consideration of variables closely related to knowledge acquisition.

The concomitant changes in curricula toward more authentic goals necessitated by any widescale implementation of performance assessment could offer special students several benefits. For example, simulations provide valuable opportunities for rehearsal prior to confronting real situations. The early and consistent focus on real-world goals gives students with disabilities specific and concrete tasks on which to focus their efforts.

In response to growing demands for more inclusion, performance assessment could confirm the practice of offering students with disabilities the same curriculum and assessment as students without disabilities. Using a variety of ca-

reer settings as frames for authentic assessment and instruction might better prepare students for their future. Among other notable features of performance assessment that could impact special students positively are its capacity to strengthen the bonds between assessment and intervention and complement educational programs that emphasize instruction/outcomes interaction (Elliott, 1992), key elements for facilitating learning. Authentic assessment even holds promise for "achieving an appropriate instructional match for every student" (Christenson, 1992, p. 296).

The possible uses and benefits of performance assessment for students with disabilities are among the critical issues presently being studied by the DID-ERIC/OSEP Special Project working group. A preliminary comment of the group suggests that performance assessment offers the potential to "appropriately and validly measure the learning achievement of students with disabilities" (Coutinho & Malouf, 1993).

General Disadvantages

Siegel (1986) categorized the disadvantages of performance assessment into four areas: (a) scoring problems; (b) expense; (c) questionable utility; and (d) inadequate long-term predictive validity. Additional problems are associated with each of these factors.

Evaluation of performance assessments can be quite subjective. Evaluation criteria should match the judgment of the performance if it had taken place in an authentic setting (Carroll & Hall, 1985). Examiner objectivity and reliability are especially problematic (Siegel, 1986). The student must know the objective of the test (what is to be done), the conditions under which the task is to be performed (use of dictionaries, timed or untimed), and the criteria to evaluate the performance (Jacobs & Chase, 1992).

Expenses may be greater than for paper-pencil tests because of the need for more examiners and their training needs (Siegel, 1986). The amount of time involved in administering, scoring, and interpreting further increases costs (Archbald, 1992). Teachers have voiced complaints about the amount of time required. Pierce and O'Malley (1992) suggested that the time factor lessens as students and teachers become more familiar with procedures, and they offer specific suggestions for facilitating the portfolio process. According to Madaus (1993), research documents that general inefficiency is a problem associated with performance assessment.

The lack of a theoretical base for performance assessment (Coutinho & Malouf, 1993) is a pervasive problem. Siegel (1986) has questioned the utility of performance measures for evaluating decision-making or judgmental ability, sug-

gesting that utility may be greater for assessing motor skills.

Performance tests have been shown to have low generalizability (Madaus & Tan, 1993). Just because students may be able to do one task correctly does not indicate that they can do others equally well. Increasing the number of different tasks on the same test is more beneficial than increasing the number of times the task is to be performed (Linn, 1993). This point brings into doubt the predictive validity of performance measures as they are currently used, a characteristic that should be paramount for them.

Questions about the technical adequacy of performance assessment have been voiced by many individuals and groups (e.g., Coutinho & Malouf, 1993; Herman, 1992; Kamphaus, 1992; Wiggins, 1992). Perhaps the two concerns cited most frequently and emphatically focus on practicality and technical characteristics.

Possible Problems for Special Students

The general disadvantages cited, particularly problems with practicality and technical adequacy, also apply to students with disabilities. In addition, Choate and Evans (1992) have raised several questions about using authentic assessment with special learners. Their questions concern the students' abilities to handle several aspects of extended tasks, appropriateness of the criteria for special learners, and legal aspects of performance assessments. The relevance of component tasks to the "real-world needs of special students" also is questioned. The complexity of the higher academic standards reflected in authentic assessment and instruction tasks also may outdistance the capabilities of some students with disabilities (Viadero, 1993).

The lack of information about performance assessment leads to several questions for extensive study (Coutinho & Malouf, 1993): What is the appropriate use of performance assessment with students with disabilities? What specific types of assessment accommodate which particular types of disability? Which tasks are meaningful for whom? What are the options for including, accommodating, or developing performance systems? The answers to these and other questions are needed to provide guidance and avoid more problems for students with disabilities.

THE FUTURE OF PERFORMANCE ASSESSMENT

Performance assessment has garnered a multitude of testimonials and endorsements and spawned a host of innovative projects. It could well be the stimulus for more systematic implementation of effective instruction principles (Christenson, 1992). The future of performance assessment in

America, however, is not all rosy. Worthen (1993) listed 12 critical issues that will determine how well performance assessment will fare in the future. Although much remains to be done—such as answering questions about technical quality, standardization, appropriateness as high-stakes assessment, and so on—he suggested that we should move onward, using performance assessment whenever feasible in low-stakes settings and profit from that use to improve it for the future. Real, practical problems of management, time, and cost of using authentic assessment in England resulted in a return to traditional paper-and-pencil testing for an entire class to take at one time, according to Madaus and Kellaghan (1993). They noted that "this retreat from direct assessment of complex performances might suggest that the life of 'authentic' testing, designed to serve a multiplicity of purposes, may not be a very long one" (p. 469).

Specific Concerns

Even though adequate funding has not been allocated to research and develop the new assessment approaches or investigate how to administer and score performance assessments, how to equate, link, and calibrate different assessments, and resolve the issues of test fairness (Ravitch, 1993), these topics must be pursued if the future of performance assessment is to be built on a sound knowledge base. Linn (1993) noted that the use of performance tests does not automatically guarantee characteristics such as validity, fairness or lack of bias toward minorities, or broad generalizability.

Reliability

The extent of agreement among raters for the quality of portfolios was found to be low to moderate for the 1992 Vermont assessment program. Reliability varied depending on variables of grade level, criteria, subject, and the like. In no situation was highly reliable scoring noticed (CRESST Reporter, 1993). Approaches to increasing interrater reliability might have to chart new courses for establishing desired reliability levels of quality information found in portfolios and possibly for other performance assessments as well. New statistical models may have to be developed.

Validity

Focusing on the character and consequences of performance assessment are new approaches to the study of validity of assessment. This expanded viewpoint, the consequences criterion, has been ignored because it is difficult to evaluate and has not been part of the traditional approach to the study of validity. Studying the consequences of assessment in relation to validity requires investigating and analyzing the complex interplay between instruction and as-

assessment, also not found in the traditional approach to validity. With new approaches to assessment come new approaches to instruction. A thorough study will have to investigate the instructional context within which assessment occurs (Linn & Baker, 1992a). The viewpoint of outcomes-based instruction is pertinent to these new ideas about the consequences criterion.

Performance assessment opens new vistas but poses new dilemmas in several fields. For example, in language assessment, appraisal personnel must understand how a second language is learned so diagnosticians do not label a student as having "language disabilities" when, in effect, the label results from the diagnostician's misunderstanding, particularly when language minority students are involved. Russell and Ortiz (1988) emphasized the value of a dialogue model of communication for assessment in providing a more accurate description of linguistic skills than counting correct responses on traditional tests. They also noted the usefulness of assessment results for instructional planning. With the use of video recorders, oral language assessment can be obtained in a variety of authentic performance contexts. These real-world approaches to assessment will influence instructional practices in the academic area of oral language as well as all academic areas. The task now is to develop appropriate standards for special education students and criteria for evaluating these new approaches to assessment.

Linkage with Standards, States, and Systems

The National Council on Education Standards and Testing (1992) pointed out that certain activities are crucial in developing a first-rate system of assessments. These activities include measurement of the national standards through a variety of assessments. The report stated that "the Council finds it essential that different assessments produce comparable results in attainment of the standards." How these results are to be linked for comparison is problematic. Linn and Baker (1993) stated that not only is linking the different assessments (from different locales and different states) to the national standards being encouraged, but also linking test results from various states to the scales used by the National Assessment of Educational Progress (NAEP). They suggested that Mislevy's (1992) report should be useful in understanding issues related to linking test results while noting that some links of substantially different assessments are not possible. The requirement for adequate linking should consider not only the statistical procedures but also the way the tests are constructed (Mislevy, 1992).

Concerns About a National Assessment System

A coalition of more than two dozen diverse educational and civil rights groups, the Campaign for Genuine Account-

ability in Education, organized by the National Center for Fair and Open Testing (FairTest), recently expressed concern about a national system of assessment. This testing watchdog group warned President Clinton and Congress that a national testing system could kill state and local assessment innovations; a national test would penalize students, such as minority and low-income students, who had been deprived of a good education. The coalition supports state and local testing initiatives, limiting the number of tests given to Chapter I students, using performance assessments, utilizing NAEP assessments only periodically and as a sample only of student achievement, and developing new ways to evaluate tests independently ("Groups Renew Campaign," 1993).

Other professionals are concerned that national assessment might not be racially and culturally equitable. To point out the need for fairness in setting standards and in assessment alike, a group of professionals, including the Educational Testing Service, the College Board, and leading education researchers, outlined these nine principles to follow as officials establish national standards and assessments ("OSEP Panel," 1993):

1. New exams should be field tested with diverse populations to insure fairness to all children;
2. Educational standards should iterate the knowledge and skills students need in real life;
3. Standards and tests for evaluating progress should specifically consider minority students;
4. Schools must have appropriate resources to implement standards and improve student learning;
5. Several assessment options should be provided to evaluate progress;
6. Guidelines for using both standards and tests should be provided to schools;
7. New standards and assessment should replace specific methods and tests;
8. Resources needed for assessment may take funds from other school projects; and
9. Teachers must be included in developing both standards and assessments. (p. 1)

Several authorities have noted the need for students with disabilities to be included in the evolving system of national standards and assessments. Viadero (1993) reported on options for inclusion proposed by James Ysseldyke and colleagues at the National Center on Educational Outcomes. Among the suggestions are setting a wide range of performance levels for the tests, offering special accommodations, and using students' IEPs to measure mastery of the standards.

Generalizability

To give adequate indications of student knowledge, performance tests have to include a large number of tasks. The reason is that the accuracy of students' responses varies from one task to another. Tests that have only one or two tasks im-

ply an incorrect measure of knowledge. It is a matter of reliability (stability) and generalizability. Shavelson noted that, in the area of science, a reasonably reliable estimate of a student's understanding requires approximately 10 tasks. These requirements will be quite a cost burden for schools and states. Nevertheless, this technical problem will have to be addressed if large-scale performance assessment is to be conducted and cross-comparisons of students, schools, and states are reported (Dietel, 1993).

Student Diversity

With the NCEST report of inclusion of all students for the national standards and assessment, one wonders about implications for students with disabilities and special needs, those with limited English proficiency, and those of diverse cultures. When using performance assessment, additional questions can be raised concerning these students' ability to participate in a fair, unbiased way. Research in performance assessment is limited, and research with diverse students is even more scarce. Although we must include *all* students in the assessment process, at the same time we must make options available to those who cannot participate fully in performance assessment. If alternative forms of assessment determine alternative modes of instruction, we also need to ask if the "new" instruction is appropriate for students with disabilities as required by law.

Performance assessment possibly may alter the concept of student diversity as we now know it. Diagnostic assessment to determine eligibility for special education services may develop a new face because the concept of "disability" will be rewritten to reflect impairments in performing authentic tasks in real-life, real-world settings only after receiving related authentic instruction.

Cost

Performance assessment will be expensive in terms of overall costs because of personnel time in managing the testing program, administering tests, creating and preparing test items, and scoring. Popham (1993) suggested that genuine matrix sampling, featuring low-proportion sampling of students and of assessment tasks, should reduce the cost of authentic assessment, making it attractive to use. His recommendation of matrix sampling would combine item sampling and student sampling, thereby reducing costs.

Technology

The nationwide interest in alternative forms of assessment has resulted in changes in published tests. These modifications will be observable in the 1994 SAT II for achievement test items. On this test, more time will be available, reducing pressure; reading comprehension passages will be

longer; and students will have to produce, rather than select, some math responses (Moses, 1991).

A number of familiar tests are or soon will be available in a format that can be administered and scored by computer ("Computerized Testing," 1993). Savings in time and expense are obvious. Computer testing has some drawbacks, though. The National Center for Fair and Open Testing (FairTest) asks if the difference might be enough to penalize some students. Most people score higher on conventional paper-and-pencil tests because strategies for responding to the test items are different on a computer than when using a pencil and paper. Without the capacity to observe and monitor behavior and attention, computer assessment could further penalize students with disabilities.

Other Measurement Issues

"The new views of assessment are based on theories of learning that are unlike those assumed by conventional test theory. . . . At the very least, the new approaches require extension of concepts that are already in flux in the measurement community" (Camp, 1993, p. 186). Camp further pointed out that these concepts relate to unified and expanded notions of validity (Cronbach, 1988; Messick, 1989a, 1989b), the concept of "systematic validity" (Fredriksen & Collins, 1989), and a redefinition of accountability (Darling-Hammond, 1991).

Using the writing portfolio as an alternative assessment, Camp (1993) commented that evaluation is outside the realm of conventional psychometrics because descriptors such as "meaningful tasks that are complex, challenging, and inherently valuable to learning [and] knowledge and skills measured in the context of complex performance . . . must be measured." Other unconventional concepts to measure are performance over time, collaboration, and demonstrated processes and strategies. Camp went on to describe portfolios as being "most at odds with traditional psychometrics" (pp. 206–207). At the same time, new approaches to measurement theory seem necessary. Snow (1993) noted that one reason for measurement concerns is that the field of psychology never developed a psychology of test design.

As new forms of alternative assessments are developed, the authors must subscribe to technical characteristics, as well as to ease of administration, scoring, and interpretation. The Code of Fair Testing Practices in Education (Fremer, Diamond, & Camara, 1989) is a statement by the Joint Committee on Testing Practices (JCTP) with members from the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. This statement is the result of a 3-year effort to provide guidelines for test use in educa-

tion. This source, along with others, must be referenced as alternative assessments are developed and used.

Critics of the new alternative assessment will search to see that it meets acceptable standards established by professional organizations. Nathan and Cascio (1986) reviewed in detail the technical and legal standards for performance assessment documented in the *Standards for Educational and Psychological Testing* (AERA/APA/NCME Joint Committee, 1985) and the *Uniform Guidelines on Employment Selection Procedures* (U.S. Equal Employment Opportunity Commission et al., 1978). Any developer of new assessments should consult the latest edition of reports of appropriate professional standards.

In reporting research-supported characteristics of performance assessment, Madaus (1993) noted that the performance modality is inefficient, costly, difficult to standardize, and time consuming because of the large numbers of students involved. Because of his "wintery prognosis" of a predominantly performance-based national assessment, Madaus suggested a matrix sampling technique as an approach to accountability rather than testing such large numbers of students.

Reassessment

As the field is thrust toward new standards and assessments, new curriculum and instruction concepts are being developed. With the current emphasis on *outcomes*, we must force ourselves to pause and reassess the *process* of instruction and the process of learning. When the ends (the new assessment) become more important than the means (instruction), the quality of instruction deteriorates (Madaus, 1993). While new instruments and methods are being developed, we must strive equally for improved instructional strategies for students with disabilities based on their learning characteristics and on our new knowledge about learning.

THE ROLE OF CURRICULUM

One of the most germane concerns about the future of performance assessment is the reciprocal relationship between assessment and curriculum. As performance assessment strategies are developed and put in place, instruction within the classroom will have to change to reflect the content and required skills of those assessments.

The more authentic the tests become, and the more they are based on real life in the real world, the more changes will be required in training programs to prepare teachers to deal with a new, authentic curriculum. It will require teachers to learn about and remain abreast of the latest findings in

cognitive psychology and new effective learning and teaching practices.

In an interview by Kirst (1991), Lorrie Shepard commented that authentic assessment will influence the way teachers teach because the focus will be on thinking, not on eliminating wrong answers, as required by multiple-choice tests. Teachers will have to move away from a task analysis approach that is rigid in its choice and sequence of what skills to teach. Thinking and skill learning must be taught hand-in-hand—yes, even before all the "prerequisite skills" are mastered—and the linking of thought and skill must begin at the preschool level. The importance of a "thinking curriculum" for all students was highlighted by Resnick and Resnick (1992), who also identified performance assessment as appropriate assessment within a thinking curriculum.

Some students with disabilities have difficulty mastering high-order thinking skills because they do not do the required thinking independently, as other students often do. Consequently, teachers will have to instruct students with disabilities in these skills *directly* through modeling, giving prompts, checking student responses, giving feedback, and presenting the material repeatedly if needed. Direct instruction for higher-order thinking skills will allow students to "build a foundation on which the student can construct additional knowledge" (Moore, Rieth, & Ebeling, 1993, p. 11).

Further modifications in curricula will result from including more students with disabilities in general education classrooms. General and special education will have to unite to redesign instructional practices "so all students learn what is and will be most important to them in their current and future lives" (York, Doyle, & Kronberg, 1992, p. 2). Teachers will have to learn how to make the curriculum authentic and meaningful for all students—those with and without disabilities.

The thrust toward performance assessment also will change how special education teachers teach. Changes will be especially important for so-called mildly disabled students if high stakes, such as high school graduation, are tied to performance-based requirements. Thus, performance assessment will create a new curriculum for general and special education alike.

This new, authentic special education curriculum must reflect what employers need from people in the workplace and what learners will need to function effectively in the community: basic skills in language, reading, writing, and arithmetic. The basics must continue to be taught, but within a framework different from that of the past. Authentic instruction will follow the principles of effective instruction to teach basic skills and their application to real-world demands within a problem-solving, higher-order thinking curriculum for students with disabilities. The emphasis will

be on solving real-world problems, with instruction that is authentic across the integrated curriculum.

The movement toward a national curriculum is suggested by the National Council on Education Standards and Testing. Its report states that "for curriculum materials and strategies to be outstanding, they must be linked to the standards and assessment strategies on which the Council is focused" (National Council on Education Standards and Testing, 1992, p. G-4). The report by the Council's task force states that special education and LEP (limited English proficiency) students alike typically have been excluded from assessments, thereby placing them "outside of accountability." "A more inclusive approach toward assessment is needed if equity concerns are to be respected" (p. F-8). This link to the national standards and assessment will include special education students.

THE ROLE OF TEACHERS

Teachers' involvement and enthusiasm are critical to the success of all forms of performance assessment. Although specifically referring to Vermont's Portfolio Project, comments by Abruscato (1993) highlight the teacher's pivotal role in all types of performance assessment. He pointed out that the system's success for the future will depend upon how the teachers accept the process. Teachers must be involved and think it is useful, important, and practical. The positive attitude required for success will require inservice education and a sense of ownership. Prospective teachers also must be provided with appropriate preparation and training activities to gain skill and confidence in the process. Zessoules and Gardner (1991) suggested that the future of performance assessment will be brighter if administrators, parents, and the community share teachers' confidence and support their efforts.

THE SPECIAL EDUCATOR'S ROLE

As advocates for the rights and needs of students with disabilities, special educators obviously have to be involved in developing and implementing performance assessment. The exact nature of that involvement is not as clear, though. The contributions of special educators to the effective use of performance assessment is a major issue targeted for study by the DID-ERIC/OSEP working group. Specifically, the group seeks to clarify appropriate roles and responsibilities for special educators to assume and the preparation and support required to fulfill the roles (Coutinho & Malouf, 1993).

Inclusive education—serving students with disabilities in general education classrooms in inclusive neighborhood schools and community settings, supported by an infusion of specially trained personnel (Council for Exceptional

Children, 1993)—will necessarily redefine special educators' roles. As inclusive instruction increases, the role of the special education teacher will shift to that of co-facilitator, working with students without and with disabilities. Working with general educators, special educators will become consultants and collaborators as they cooperatively assess, or co-assess, their joint charges (Choate, in press).

As standards are set for students to achieve, special educators must see to it that students with disabilities are considered. These standards should be in all appropriate areas of disability so all students with disabilities have a framework for reaching established objectives and goals without losing sight of the appropriateness of the educational placement, instruction, and services for the individual within a special education milieu. Special education standards for the 1990s could easily be based on Aschbacher's work (News & Governance, 1993). She described contextual standards (what a person knows), procedural standards (what a person is able to do), performance standards (how well students are expected to perform when assessed), and delivery standards (quality of educational opportunities provided to enable attainment of other standards).

Special educators may have given educational reformers mixed signals on whether and how to include students with disabilities in performance assessment programs (Viadero, 1993). Similarly, special educators have not established or assumed clearly defined roles. Considering all the unanswered questions and the limited role of special educators in the process thus far, a number of actions seem appropriate, perhaps even imperative.

AN ACTION AGENDA FOR SPECIAL EDUCATORS

Special educators must take action to ensure that appropriate accommodations for students with disabilities are built into the changing assessment practices and resulting modifications in curricula. In reviewing the necessary framework for action, the questions and answers structuring this discussion are:

1. *What is performance assessment?* Performance assessment examines what students can do with knowledge. Authentic performance assessment evaluates students' achievement of realistic tasks set in real-life contexts. Portfolio assessment involves appraising the evidence or permanent products of performance assessment tasks, which may or may not be authentic.
2. *How is it currently used?* Performance assessment currently is used to assess competency, through commercial tests, in various projects, and in some statewide assessment programs, a few of which in-

clude students with disabilities in the process. Aside from scattered inclusion of special education students in state assessments and its use in vocational education, where performance assessment is a tradition, limited systematic implementation for students with disabilities is reported.

3. *What is its value to special education?* The major advantage of performance assessment is its direct connection with instruction. In addition, it avoids some of the criticisms of traditional testing. The possible benefits for special students are yet to be realized because of limited implementation, but prospects exist for appropriate instructional planning. General disadvantages include drawbacks in scoring, expense, questionable utility, and inadequate technical features. An additional concern for students with disabilities is practicality.
4. *What is its future?* Performance assessment seems to be gaining steam. Specific concerns about the performance assessment process itself must be addressed, and issues related to the role of curricula, teachers, and special educators must be resolved. Its future use and value to students with disabilities also hinges on the answer to the final question below, and on the resulting actions.
5. *What should special educators do to best accommodate students with disabilities?* Currently, special education's role is ill defined, and its participation thus far has been limited. Special educators seem to be aware of performance assessment but not active in its evolution. Clearly, special educators must become involved, and the Performance Assessment Working Group, OSEP/ERIC Special Project with the Council for Exceptional Children, has begun the process.

Several actions are required. Together, they constitute an "action agenda" for special educators. Specifically, the recommended agenda for action is this:

- *Get involved.* The main task to accomplish is involvement—immediate and active involvement. We must be involved intimately in developing, implementing, and researching performance assessment. Thus, our first and primary role is defined: Special educators will be *active participants* in performance assessment. Better yet, special educators will step forward to assume leadership roles in the evolution of the performance assessment movement.
- *Directly address the issues.* The first issue to address is the need to develop authentic curricula so that what is tested is what is taught. If tests are to become oriented

more to real life in the real world, what is taught to students with and without disabilities must be oriented the same way. The second issue is the need to appropriately include students with disabilities in the performance curriculum and assessment process. We must actively seek answers to the real and potential problems performance assessment poses for students with disabilities.

- *Monitor progress.* As a national system of assessment is developed, we must serve as watchdogs to guarantee that program evaluations of students with learning disabilities (and other disabilities) are not overlooked and that assessments to measure educational progress and accountability are appropriate (National Joint Committee on Learning Disabilities, 1993). Similarly, as performance assessment is implemented locally, regionally, and statewide, policies and procedures must be checked carefully to confirm appropriate practices for special students.
- *Become and remain informed.* Know the resources for tracking performance assessment activity, the opportunities for active participation, and the avenues for contributing to the knowledge base. Peruse the reference list of this article for resources. Maintain a list and regularly consult sources such as those cited in Table 4. Although these are not comprehensive, they provide a good beginning for expanding knowledge and offer contacts for becoming actively involved.
- *Contribute to the knowledge base.* Seek opportunities to observe and participate in assessment and instructional practices such as authentic projects, portfolios, extended assignments, collaborative research, peer-mediated judgments and critiques. Then share your experiences, preferably in writing, so others may also learn from them. Explore ways to improve the performance assessment process. Communicate with key individuals who are associated with the resources such as the ones listed in Table 4. Routinely exchange information with colleagues. Develop or participate in research studies of performance assessment. Most important, contribute your findings to improve the performance assessment knowledge base.

The entire field of performance assessment is developing and evolving and thus is subject to refinement or even reformulation (Herman, Aschbacher, & Winters, 1992). Special educators must take appropriate actions—now, during the formative stages—to confirm the integrity of the practices and prospects of authentic performance assessment and authentic curricula for students with disabilities in particular and for all students in general.

TABLE 4
Sources of Information About Performance Assessment

Education Agencies	<p>Local Education Agencies General education, evaluation services, and special education diagnostic services may offer pertinent information and activities</p> <p>State Education Agencies State departments of education, state programming and/or evaluation services, special education diagnostic services may offer information/activities</p>
Networks	<p>Authentic Assessment Network, Association for Supervision and Curriculum Development Kathleen Busick, Pacific Regional Educational Lab, 1164 Bishop St., Suite 1409, Honolulu, HI 96813 (808/532-1900) or Judith Backes, Carroll County Public Schools, 55 N. Court St., Westminster, MD 21157 (410/876-2208)</p> <p>Performance Assessment Network-Council for Educational Diagnostic Services James A. Poteet, Professor, Teachers College, Ball State University, Muncie, IN (317/285-5707); InterNET 00JAPoteet@BSUVC.BSU.EDU</p>
Newsletters	<p>Authentic Assessment Network News Kathleen Busick, Pacific Regional Educational Lab, 1164 Bishop St., Suite 1409, Honolulu, HI 96813 (808/532-1900) or Judith Backes, Carroll County Public Schools, 55 N. Court St., Westminster, MD 21157 (410/876-2208)</p> <p>Portfolio Assessment Newsletter Northwest Evaluation Association; c/o Allan Olson, 5 Centerpointe Dr., Suite 100, Lake Oswego, OR 97035</p> <p>Portfolio News W. Cooper & J. Daves, c/o San Dieguito Union High School District, 710 Encinitas Boulevard, Encinitas, CA 92024</p>
Special Projects	<p>Center for Research on Evaluation, Standards, and Student Testing UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, CA 90024-1522.</p> <p>Performance Assessment Working Group, ERIC/OSEP Special Project of Council for Exceptional Children (CEC) 703/620-3660</p>

REFERENCES

- Abruscato, J. (1993). Early results and tentative implications from the Vermont portfolio project. *Phi Delta Kappan*, 74(6), 474-477.
- AERA/APA/NCME Joint Committee. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychological Association. (1991). A national exam in education? *Psychological Science Agenda*, 4(3), 2.
- Archbald, D. A. (1992). Authentic assessment: Principles, practices, and issues. *School Psychology Quarterly*, 6(4), 279-293.
- Archbald, D. A., & Newmann, F. M. (1988). *Beyond standardized testing*. Reston, VA: National Association of Secondary School Principals.
- Camp, R. (1993). The place of portfolios in our changing views of writing assessment. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 183-212). Hillsdale, NJ: Lawrence Erlbaum.
- Carroll, B. J., & Hall, P. J. (1985). *Make your own language tests*. Oxford: Pergamon.
- Choate, J. S. (in press). Co-assessment of special learners: A call for special and general education to unite. *Preventing School Failure*.
- Choate, J. S., Enright, B. E., Miller, L. J., Poteet, J. A., & Rakes, T. A. (1992). *Curriculum-based assessment and programming* (2nd ed.). Boston: Allyn and Bacon.
- Choate, J. S., & Evans, S. S. (1992). Authentic assessment of special learners: Problem or promise? *Preventing School Failure*, 37(1), 6-9.
- Christenson, S. L. (1992). Authentic assessment: Straw man or prescription for progress? *School Psychology Quarterly*, 6(4), 294-299.
- Computerized testing may skew results, FairTest says. (1993). *Education Daily*, January 21, p. 7.
- Council for Exceptional Children (1993). CEC policy on inclusive schools and community settings. *Teaching Exceptional Children*, 25(4), supplement.
- Coutinho, M., & Malouf, D. (1993). Performance assessment and children with disabilities: Issues and possibilities. *Teaching Exceptional Children*, 25(4), 62-67.
- CRESST Reporter. (1993, Winter). *The CRESST Line: Newsletter of National Center for Research on Evaluation, Standards, and Student Testing*, p. 8.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Waner & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Darling-Hammond, L. (1991). Measuring schools is not the same as improving them. In *Voices from the field: 30 expert opinions on "America 2000," The Bush administration strategy to "reinvent" America's schools*. Washington, DC: William T. Grant Foundation

Commission on Work, Family, and Citizenship and Institute for Educational Leadership.

- Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan*, 74(10), 752–761.
- Dietel, R. (1992, Fall). Work in progress: Portfolios as worthwhile burdens? *CRESST Line Newsletter of National Center for Research and Evaluation, Standards, and Student Testing*. UCLA, Graduate School of Education, Los Angeles: pp. 1–5.
- Dietel, R. (1993, Winter). Work in progress: Re-assessing group assessment. *CRESST Line Newsletter of National Center for Research and Evaluation, Standards, and Student Testing*. UCLA, Graduate School of Education, Los Angeles, pp. 6, 8.
- Educational Testing Service. (1990). *The educational reform decade*. Princeton, NJ: Author.
- Elliott, S. N. (1992). Authentic assessment: An introduction to a neo-behavioral approach to classroom assessment. *School Psychology Quarterly*, 6(4), 273–278.
- Feuer, M. J., & Fulton, K. (1993). The many faces of performance assessment. *Phi Delta Kappan*, 74(6), 478–479.
- Frederiksen, J. R. & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18(1), 27–32.
- Fremer, J., Diamond, E. E., & Camara, W. J. (1989). Developing a code of fair testing practices in education. *American Psychologist*, 44(8), 1062–1067.
- Gitomer, D. H. (1993). Performance assessment and educational measurement. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 241–263). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Grace, C. (1992). The portfolio and its use: Developmentally appropriate assessment of young children. (ERIC Document Reproduction Service No. ED 351 150)
- Grady, E. (1992). *The portfolio approach to assessment* (Fastback 341). Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Gresham, F. M. (1992). Alternative psychometrics for authentic assessment. *School Psychology Quarterly*, 6(4), 305–309.
- Groups renew campaign against national assessment. (1993). *Education Daily*. March 5, p. 3.
- Hambleton, R., & Murphy, E. (1991). A psychometric perspective on authentic measurement (Report No. L 214). Amherst, MA: Laboratory of Psychometric and Evaluative Research. (ERIC Document Reproduction Service No. ED 334 265 TM 016 890)
- Hambleton, R., & Oakland, T. (1993). Organization profile: International Test Commission. *Psychology International*, 4(3), 8.
- Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, what, and whithers. *Phi Delta Kappan*, 70(9), 683–687.
- Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley: University of California Press.
- Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership*, 49(8), 74–78.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision & Curriculum Development.
- Jacobs, L. C., & Chase, C. I. (1992). *Developing and using tests effectively*. San Francisco: Jossey-Bass.
- Kamphaus, R. W. (1992). Authentic assessment and content validity. *School Psychology Quarterly*, 6(4), 300–304.
- Kaya, E. (1974). Hostra-Manhassent Project competency evaluation. *Performance-based Teacher Education*, 3(2), 9–12.
- Kirst, M. (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher*, 20(2), 21–23.
- Koretz, D. M., Madaus, G. F., Haertel, E., & Beaton, A. E. (1992). *National educational standards and testing: A response to the recommendations of the National Council on Education Standards and Testing*. Santa Monica, CA: Rand.
- Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation & Policy Analysis*, 15(1), 1–16.
- Linn, R. L., & Baker, E., L. (1992, Summer). From the Directors: Consequences of performance assessment. *CRESST Line: Newsletter of National Center for Research on Evaluation Standards, and Student Testing*. UCLA, Graduate School of Education, Los Angeles, pp 2, 7.
- Linn, R. L., & Baker, E., L. (1992, Fall). From the Directors: Portfolios and accountability. *CRESST Line: Newsletter of National Center for Research on Evaluation, Standards, and Student Testing*. UCLA, Graduate School of Education, Los Angeles, pp 1, 8–9.
- Linn, R. L., & Baker, E. L. (1993, Summer). Comparing results from disparate assessments. *CRESST Line: Newsletter of National Center for Research on Evaluation, Standards, and Student Testing*. UCLA, Graduate School of Education, Los Angeles, pp 1–2.
- Livingston, C., Castle, S., & Nations, J. (1989). Testing and curriculum reform: One school's experience. *Educational Leadership*, 46(7), 23–25.
- Madaus, G. F. (1993). A national testing system: Manna from above? An historical/technological perspective. *Educational Assessment*, 1(1), 9–26.
- Madaus, G. F., & Kellaghan, T. (1993). The British experience with “authentic” testing. *Phi Delta Kappan*, 74(6), 458–469.
- Madaus, G. F., & Tan, A. G. (1993). The growth of assessment. In G. Cawelti (Ed.), *Challenges and achievements of American education*, 1993 Yearbook (pp. 53–79). Alexandria, VA: Association for Supervision and Curriculum Development.
- Messick, S. (1989a). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(1), 5–14.
- Messick, S. (1989b). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: Macmillan.
- Meyer, C. A. (1992). What's the difference between authentic and performance assessment? *Educational Leadership*, 49(8), 39–40.
- Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: Educational Testing Service.
- Moore, P. R., Rieth, H., & Ebeling, M. (1993). Considerations in teaching higher order thinking skills to students with mild disabilities. *Focus on Exceptional Children*, 25(7), 1–12.
- Moses, S. (1990, August). Panel says tests shut too many out. *APA Monitor*, p. 28.
- Moses, S. (1991, January). Major revision of SAT goes into effect in 1994. *APA Monitor*, p. 35.
- Moses–Zirkes, S. (1992, November). Students with disabilities are left out of testing data. *APA Monitor*, p. 45.
- NAGB adopts policy on future of NAEP. (1993). *Education Daily*, March 9, p. 4.
- Nathan, G. R., & Cascio, W. F. (1986). Introduction, technical, and legal standards. In R. A. Beck (Ed.), *Performance assessment: Methods & applications* (pp. 1–50). Baltimore: The Johns Hopkins University Press.
- National Center for Fair and Open Testing (FairTest) (1992). *What is authentic evaluation?* Cambridge, MA: Author. (ERIC Document Reproduction Service No. ED 352 373)
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America*. Chestnut Hill, MA: Author.

FOCUS ON EXCEPTIONAL children

- National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Washington, DC: U.S. Government Printing Office.
- National Joint Committee on Learning Disabilities. (1992). School reform: Opportunities for excellence and equity for individuals with learning disabilities. *Journal of Learning Disabilities*, 25(5), 276-280.
- News and Governance. (1993, Winter). *CRESST Line: Newsletter of National Center for Research on Evaluation, Standards, and Student Testing*. UCLA, Graduate School of Education, Los Angeles, p. 3.
- Neil, D. M., & Medina, N. J. (1989). Standardized testing: Harmful to educational health. *Phi Kappa Kappan*, 70(9), 68-69.
- Nolet, V. (1992). Classroom-based measurement and portfolio assessment. *Diagnostique*, 18(1), 5-26.
- Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions* (OTA-SET-519). Washington, DC: U.S. Government Printing Office (ED 340-770).
- OSEP panel wants special ed concerns addressed in assessments. (1993). *Education Daily*, March 30, p. 8.
- Pierce, L. V., & O'Malley, J. M. (1992). *Performance and portfolio assessment for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.
- Popham, W. J. (1993). Circumventing the high costs of authentic assessment. *Phi Delta Kappan*, 74(6), 470-473.
- Ravitch, D. (1993). Launching a revolution in standards and assessments. *Phi Delta Kappan*, 74(10), 767, 772.
- Record number of students tool PSAT/NMSQT last year. (1993). *Education Daily*, March 5, p. 4.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.) *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 38-75). Boston: Kluwer Academic Publishers.
- Roeber, E., Bond, L., & van der Ploeg, A. (1993). *State student assessment program data base*. Washington, DC: Council of Chief State School Officers/North Central Regional Education Laboratory.
- Russell, N. L., & Ortiz, A. A. (1988). Assessment and instruction within a dialogue model of communication: Part 1. *Bilingual special education newsletter*, 8 (Fall issue). Austin, TX: University of Texas.
- Siegel, A. I. (1986). Performance tests. In R. A. Beck (Ed.), *Performance assessment* (121-142). Baltimore: Johns Hopkins University Press.
- Sitlington, P. L. (1979). Vocational assessment and training of the handicapped. *Focus on Exceptional Children*, 12(4), 1-11.
- Shohamy, C. (1992). New modes of assessment: The connection between testing and learning. In E. Shohamy & A. R. Walton (Eds.), *Language assessment for feedback: Testing and other strategies* (pp. 7-28). Dubuque, IA: Kendall/Hunt.
- Slater, S. (1980). Introduction to performance testing. In J. Spierer (Ed.) *Performance testing: Issues facing vocational education*. Columbus, OH: National Center for Research in Vocational Education.
- Snow, R. E. (1993). Construct validity and constructed-response tests. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum Associates.
- U.S. Department of Education. (1991). *America 2000: An education strategy*. Washington, DC: Author.
- U.S. Equal Employment Opportunity Commission, U.S. Civil Service Commission, U.S. Department of Labor, & U.S. Department of Justice. (1978). Uniform guidelines on employment selection procedures. *Federal Register*, 43(166), 38290-38309.
- Viadero, D. (1993). The rhetoric and reality of high academic standards. *Education Week*, 12(36), 1, 16-17.
- Wesche, M. B. (1992). Performance testing for work-related second language assessment. In E. Shohamy & A. R. Walton (Eds.), *Language assessment or feedback: Testing and other strategies* (pp. 103-122). Dubuque, IA: Kendall/Hunt.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(6), 41-47.
- Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership*, 49(8), 26-33.
- Worthen, B. R. (1993). Critical issues that will determine the future of alternative assessment. *Phi Delta Kappan*, 74(6), 444-454.
- York, J., Doyle, M. B., & Kronberg, R. (1992). A curriculum development process for inclusive classrooms. *Focus on Exceptional Children*, 25(4), 1-16.
- Zessoules, R., & Gardner, H. (1991). Authentic assessment: Beyond the buzzword and into the classroom. In V. Perrone (Ed.), *Expanding students assessment* (pp 47-71). Alexandria, VA: Association for Supervision & Curriculum Development.

PERMISSIONS AND COPYRIGHT

All rights are reserved. No part of this publication may be reproduced, photocopied, faxed, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, recording or otherwise, without the prior written permission of the publisher.

Back issues are available for sale. Reproduction requires permission and payment of fees. It is illegal and a violation of Federal copyright law to reproduce this publication without permission. Direct all inquiries to the permissions editor.