DOCUMENT RESUME

ED 375 568 EC 303 405

AUTHOR McLaughlin, Margaret J.; Warren, Sandra Hopfengardner

TITLE Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems. Performance Assessment: CEC Mini-Library.

INSTITUTION Council for Exceptional Children, Reston, Va.; ERIC Clearinghouse on Disabilities and Gifted Education, Reston, VA. ERIC/OSEP Special Project on Interagency Information Dissemination.

SPONS AGENCY Special Education Programs (ED/OSERS), Washington, DC.

REPORT NO ISBN-0-86586-250-8

PUB DATE 94

NOTE 53p.

AVAILABLE FROM Council for Exceptional Children, 1920 Association Dr., Reston, VA 22091-1589 ($8.90 non-members, $6.25 members; Stock No. P5061; $32 set, non-members, $22.50 set, members; Stock No. P5062).

PUB TYPE Information Analyses (070) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS Academic Standards; *Accountability; *Disabilities; Educational Assessment; Educational Objectives; Elementary Secondary Education; *Evaluation Methods; Outcomes of Education; School Districts; State Programs; *Student Educational Objectives; *Student Evaluation

IDENTIFIERS *Performance Based Evaluation; *Testing Accommodations (Disabilities)

ABSTRACT

As part of the educational reform movement, states and local districts have implemented outcomes-based accountability for student outcomes. Outcomes-based accountability systems are using results of both traditional assessments, such as norm-referenced tests, and authentic performance assessments in an effort to measure broad domains of student knowledge. Issues in using performance assessments in large-scale on-demand assessment programs include the costs associated with administration and reliability in scoring. The level of rigor applied to assessment of outcomes for nondisabled students compared to students who are receiving special education services is being considered. Issues in using performance assessments for students with disabilities include defining the outcomes to be assessed, developing performance standards, developing assessment accommodations, and scoring. Five performance assessment programs that have taken various approaches to the inclusion of students with disabilities are described; these include Kentucky; Maryland; Vermont; Littleton, Colorado; and Arlington Heights, Illinois. Interviews with representatives of these programs indicated that students with disabilities could be exempted from participation in the assessment program (except in Kentucky) and that decisions to exempt were typically made by Individualized Education Program teams. (Contains 29 references.) (JDD)

PERFORMANCE ASSESSMENT

Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems

Margaret J. McLaughlin and
Sandra Hopfengardner Warren

Funded by the Office of Special Education Programs, U.S. Department of Education

The Council for Exceptional Children

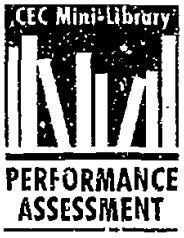
CEC: Leading the Way
The Council for Exceptional Children (CEC) is the largest professional organization internationally committed to improving educational outcomes for individuals with exceptionalities. CEC accomplishes its worldwide mission on behalf of educators and others working with children with exceptionalities by advocating for appropriate government policies; setting professional standards; providing continuing professional development; and assisting professionals to obtain conditions and resources necessary for effective professional practice.

CEC: The Unifying Force of a Diverse Field
The Council for Exceptional Children, a private nonprofit membership organization, was established in 1922. CEC is an active network of 59 State/Provincial Federations, 900 Chapters, 17 Specialized Divisions, and 275 Subdivisions with reach in over 40 countries.

The CEC Information Center:
International Resource for Topics in Special and Gifted Education
The Council for Exceptional Children is a major publisher of special education literature and produces a comprehensive catalog semiannually. Journals such as *TEACHING Exceptional Children* (published quarterly) and *Exceptional Children* (published 6 times per year) reach over 100,000 readers and provide a wealth of information on the latest teaching strategies, research, resources, and special education news. *CEC Today*, the Council's monthly communications vehicle, provides the latest on Association issues, events, and advocacy efforts.

To help those in the field better understand the value of alternative assessment practices, CEC is proud to release *Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems.*

The Council for Exceptional Children
1920 Association Drive • Reston, Virginia 22091
703/620-3660 (Voice/TDD)
703/264-9494 (Fax)



Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems

Margaret J. McLaughlin and
Sandra Hopfengardner Warren

Published by
The Council for Exceptional Children

A Product of the
ERIC/OSEP Special Project
The ERIC Clearinghouse on
Disabilities and Gifted Education

Library of Congress Cataloging-in-Publication Data

McLaughlin, Margaret J.
Performance assessment and students with disabilities : usage in outcomes-based accountability systems / Margaret J. McLaughlin and Sandra Hopfengardner Warren.
p. cm.—(CEC mini-library performance assessment)
"A product of the ERIC/OSEP Special Project, the ERIC Clearinghouse on Disabilities and Gifted Education."
Includes bibliographical references.
ISBN 0-86586-250-8
1. Educational tests and measurements—United States. 2. Handicapped students—United States—Rating of. 3. Competency based education—United States. 4. Educational accountability--United States. I. Warren, Sandra Hopfengardner. II. Title. III. Series.
LB3051.M46249 1994
371.2'7—dc20 94-26214
CIP

ISBN 0-86586-250-8

A product of the ERIC/OSEP Special Project, the ERIC Clearinghouse on Disabilities and Gifted Education

Published in 1994 by The Council for Exceptional Children, 1920 Association Drive, Reston, Virginia 22091-1589

Stock No. P5061

This publication was prepared with funding from the U.S. Department of Education, Office of Special Education Programs, contract no. RR93002005. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication the manuscript was submitted for critical review and determination of professional competence. This publication has met such standards. Points of view, however, do not necessarily represent the official view or opinions of either The Council for Exceptional Children or the Department of Education.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Foreword

CEC's policy on inclusive schools and community settings invites all educators, other professionals, and family members to work together to create early intervention, educational, and vocational programs and experiences that are collegial, inclusive, and responsive to the diversity of children, youth, and young adults. Policymakers at the highest levels of state/provincial and local government, as well as school administration, also must support inclusive principles in the educational reforms they espouse.

One area in which the inclusion of students with disabilities is critical is the development and use of new forms of assessment. This is especially true when assessment becomes a tool by which local school districts, states, and our nation show accountability for the education of students.

As multidimensional instruments that can cross curriculum areas, performance assessments have the potential to be powerful instructional tools as well as tools for accountability. As this new technology is applied in creating new assessment instruments, students with disabilities must be considered during the design of the assessment, administration, scoring, and reporting of results.

CEC is proud to contribute this Mini-Library to the literature on performance assessment, and in so doing to foster the appropiate inclusion of students with disabilities in this emerging technology for instruction and accountability.

Preface

Performance assessment, authentic assessment, portfolio assessment—these are the watchwords of a new movement in educational testing. Its advocates say this movement is taking us beyond the era when the number 2 pencil was seen as an instrument of divine revelation. Its critics say it is just another educational bandwagon carrying a load of untested techniques and unrealistic expectations.

Despite the criticisms and reservations that are sometimes expressed, these new approaches are being implemented in a growing number of large-scale assessment programs at federal, state, and district levels. They are also finding their way into small-scale use at school and classroom levels.

What about students with disabilities? Are the new assessment techniques more valid than conventional assessment techniques for these students? Are the techniques reliable and technically sound? Will they help or hinder the inclusion of students with disabilities in large-scale assessment programs? Can classroom teachers use the techniques to assess student learning and possibly enrich the classroom curriculum?

The following fictional vignettes illustrate some of these issues.

Vignette 1

The State of Yorksylvania developed educational standards and a statewide system of student assessments to monitor progress in achieving the standards. The use of standardized multiple-choice tests was rejected because these tests were thought to trivialize education. It was feared that teachers would "teach down" to the tests rather than "teach up" to the standards. So, committees of teachers, parents, and employers were formed to translate the standards into "authentic" performance assessments. The resulting assessment system was called the Yorksylvania Performance Inventory (YPI).

Once a year, students from every school in the state were administered the YPI, which consisted of several assessments, each of which required up to 3 days to complete. Students worked, sometimes individually and sometimes in small groups, on tests involving complex, high-level tasks that crossed curriculum areas. In one task, students individually did research and answered essay questions interrelating the geography, wildlife, and history of their state. In another task, students worked in groups to design a car powered by fermentation. Schools were provided with practice activities and curriculum guides to encourage the infusion of performance assessment activities into the school curriculum.

The state policy allowed special education students to be included in the YPI, excluded, or provided with special modifications, depending on their individual needs as indicated in their individualized education programs. Initially, most special education teachers supported the YPI because they felt it eliminated some artificial barriers (reading, test-taking skills, etc.) that put their students at a disadvantage on other types of tests. However, there were some questions and issues, such as the following:

- Some of the YPI tasks involved *a lot* of reading, more than was found on previous types of tests.
- Special education teachers sometimes felt pressured to exclude their students from testing in order to increase the school's scores.
- Special education students sometimes experienced extreme frustration in the YPI assessments, many of which bore no resemblance to these students' other schoolwork.
- Some parents of special education students questioned whether the standards were really applicable to their children and whether the YPI was diverting instruction from more relevant and important topics.

Vignette 2

A teacher named Pat had students at a wide range of functioning levels, including a number of mainstreamed students receiving special education services. Pat was always on the lookout for new ideas and approaches. Pat began reading articles and attending conferences on new assessment approaches termed *portfolio assessment, authentic assessment, per-*

formance assessment, and *alternative assessment*. These approaches seemed to make a lot of sense, and Pat decided to try them out. One of the first approaches Pat tried was authentic assessment. Rather than simply testing students on their rote learning of skills and content, Pat began to look for ways to use realistic, complex activities to test whether the students could actually apply what they learned. For example, Pat combined writing, spelling, science, and career skills into an activity in which students wrote letters of application for jobs as physicists, biologists, or chemists. Pat particularly valued activities that engaged students in solving interesting problems. For example, after a unit on optics, Pat assigned students to draw a diagram explaining why mirrors reverse an image from left to right but not from top to bottom. The students grappled with that problem for several days.

Pat liked the holistic scoring procedures developed in these new assessment approaches. Rather than simply marking a response correct or incorrect, Pat scored student work on a number of dimensions (e.g., analysis of the problem, clarity of communication) according to meaningful quality criteria. The development of authentic performance tasks and scoring procedures helped Pat clarify the most important learning outcomes.

Pat also liked the idea of portfolio assessment, in which students could select and collect "best pieces" to demonstrate their learning and achievement during the year. Student self-evaluation became a valued part of this process.

In all, Pat was very pleased with these new assessment approaches and intended to continue using them. Instruction became more activity based and more focused on real-world uses of the material. There were, however, some issues that Pat began to think about:

- Students with deficits in certain academic areas, notably writing, were at a real disadvantage. It was sometimes hard to determine whether an inadequate response resulted from poor writing skills, poor mastery of the content, poor problem-solving skills, lack of creativity, or some combination of these factors. Pat considered allowing some students to tape record their responses, but decided not to. Wasn't writing itself an authentic task required in the real world?

- Pat wasn't sure how to use the information provided by these tests to plan additional instruction, particularly if a student was having difficulty.
- Pat wondered how to tell whether or not an activity was in fact authentic, especially for students whose adult lives would be very different from Pat's own.

In 1992, the Division of Innovation and Development (DID) in the U.S. Department of Education's Office of Special Education Programs and the ERIC/OSEP Special Project of The Council for Exceptional Children formed a Performance Assessment Working Group to discuss issues such as these. The term *performance assessment* was adopted as a general designation for the range of approaches that include performance assessment, authentic assessment, alternative assessment, and portfolio assessment.

Performance assessment was defined has having the following characteristics:

1. *The student is required to create an answer or a product rather than simply fill in a blank, select a correct answer from a list, or decide whether a statement is true or false.*
2. *The tasks are intended to be "authentic."* The conventional approach to test development involves selecting items that represent curricular areas or theoretical constructs, and that have desired technical characteristics (e.g. they correlated with other similar items, they discriminated between groups, etc.). Authentic tasks, on the other hand, are selected because they are "valued in their own right"[1] rather than being "proxies or estimators of actual learning goals."[2]

The Performance Assessment Working Group produced this series of four Mini-Library books on various topics related to performance assessment and students with disabilities. In *National and State Perspectives on Performance Assessment and Students with Disabilities*, Martha Thurlow discusses trends in the use of performance assessment in large-scale testing programs. In *Performance Assessment and Students with Disabilities: Usage in Outcomes-Based Accountability Systems*, Margaret McLaughlin and Sandra Hopfengardner Warren describe the experi-

[1]R. L. Linn, E. L. Baker, & S. B. Dunbar. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.
[2]M. W. Kirst. (1991). Interview on assessment issues with Lorrie Shepard. *Educational Researcher. 20*(2), 21–23, 27.

ences of state and local school districts in implementing performance assessment. In *Creating Meaningful Performance Assessments: Fundamental Concepts*, Stephen Elliott discusses some of the key technical issues involved in the use of performance assessment. And, in *Connecting Performance Assessment to Instruction*, Lynn Fuchs discusses the classroom use of performance assessment by teachers.

Martha J. Coutinho
University of Central Florida

David B. Malouf
U.S. Office of Special Education Programs

August, 1994

Members of the Performance Assessment Work Group

Joan Baron, Performance Assessment Collaborative for Education, Harvard University

Joyce Choate, Council for Exceptional Diagnostic Services, Northeast Louisiana University

Lorraine Costella, Frederick County, Maryland, Public Schools

Martha Coutinho, Division of Innovation and Development, U.S. Office of Special Education Programs

Stephen Elliott, University of Wisconsin-Madison

Lynn Fuchs, Vanderbilt University

John Haigh, Maryland State Department of Education

Larry Irvin, University of Oregon

Robert Linn, Center for Research on Evaluation, Standards, and Student Testing

Lynn Malarz, Association for Supervision and Curriculum Development

David Malouf, Division of Innovation and Development, U.S. Office of Special Education Programs

Margaret McLaughlin, Center for Policy Options in Special Education, University of Maryland

Trina Osher, National Association of State Directors of Special Education

James Poteet, Council for Educational Diagnostic Services, Ball State University

Clay Starlin, Western Regional Resource Center, University of Oregon

Martha Thurlow, National Center on Educational Outcomes for Students with Disabilities, University of Minnesota

Sandra Hopfengardner Warren, Center for Policy Options in Special Education, University of Maryland

Note: Members' affiliations may have changed since the work group was formed.

About the Authors

Margaret J. McLaughlin and **Sandra Hopfengardner Warren** are both with the Institute for the Study of Exceptional Children and Youth at the University of Maryland. For the Institute, they recently completed a federal contract to operate the Center for Policy Options in Special Education. The Center was a 3-year effort that addressed critical issues and policy options related to school restructuring and students with disabilities, including issues related to outcomes-based accountability.

"We must make children with disabilities an integral part of the educational enterprise," says Dr. McLaughlin. "We must also make parents understand that children with disabilities can share common educational goals with general education students. In the context of broad goals for education, the things we value for other kids are the things we value for these kids, too. Together, we must step back and look at what today's children will need to know when they are adults. No one person can define that."

Dr. McLaughlin's interest in special education developed in her college years at the University of Denver, where she majored in elementary education and special education for students with emotional disabilities. After 2 years working at an institution for children with psychiatric disorders, she returned to Colorado for her master's degree. She then worked in public schools in Newport News and Colonial Heights, Virginia, as teacher of self-contained and resource classes that included students with mild mental retardation, emotional disturbance, and learning disabilities.

She had always been interested in policy systems and the overall view of special education, believing that as a practical matter, special and general education should be part of the same system while still providing individualized education to meet the special needs of students. To further pursue this interest, she attended the University of Virginia for her doctorate, then served an internship at what was then the U.S. Bureau of Education for the Handicapped. She moved on to the University of Washington and then to the University of Maryland.

Among the projects she is involved with at the University of Maryland is the Center for Urban Special Education, a school/university partnership between Baltimore City Public Schools and the University of Maryland. Additionally, she is directing a new effort, under subcontract to the Maryland State Department of Education, that will develop alternative performance assessments for students with severe cognitive disabilities. These assessments will be part of the Maryland State Assessment System. She has written extensively on special education policy and teaches graduate courses in disability policy at the University of Maryland.

Dr. Sandra Hopfengardner Warren sees policy as a vehicle to bring about changes in practice. She believes that there is a need for systemic organizational reform to provide more effective educational experiences for youngsters. Her dedication to this belief developed when, as an undergraduate, she volunteered at an institution for individuals with developmental disabilities. In addition to the volunteer work, she recruited other students to serve as volunteers. She and the other students were appalled by the deplorable conditions in which the residents lived. To Dr. Warren, "it became clear that the path to a better life for these residents was to get them out into the community and for the community to get involved in their lives—be it through education, employment, or other community involvement."

Her graduate work was in public administration with an emphasis on organizational development, and following graduation she became Executive Director of United Cerebral Palsy of Washington, DC. Among other efforts, she began a collaborative effort with the local Head Start Agency to provide inclusive education for preschoolers. She later served 5 years as a court monitor to help close the only state-operated institution for individuals with developmental disabilities in Washington, DC, and to develop a community support system. "This was an exciting time for me to study the relationship between advocacy and policy," she says. "Now, kids and adults with disabilities are living in the community next door to nondisabled students, employers, policy makers, and other adults. With this proximity, the doors to communities are opened for all people—including schools for all students."

Her interest in the power of policy led her to pursue her Ph.D. in disability policy, and then to continue this focus in her work at the University of Maryland. "The Center for Policy Options in Special Education offered an excellent opportunity to study the issues in education for these students," she says. "With the advent of systemic reform and restructuring, we are focusing our attention on issues of accountability. Performance assessments appear to offer an avenue for including students with disabilities, including those with severe disabilities, in the accountability system." Dr. Warren is excited about her current work at

the university developing performance assessments for the State of Maryland so that students with severe cognitive disabilities are included in the accountability system.

Margaret McLaughlin and Sandra Warren recently authored *Issues and Options in Restructuring Schools and Special Education Programs*, another CEC publication.

Contents

1. Introduction

What skills do we expect young Americans to have when they complete their education? How do we assess those skills, and how will schools be held accountable for student attainment of those skills? These are among the most commonly asked questions in today's educational reform movement. While the answers vary from state to state, a common theme is the increasing use of performance assessments and other authentic assessments to gauge student progress toward the desired student outcomes.

As part of the educational reform movement, U.S. educators and policymakers have endorsed the concept of accountability for student outcomes (Brauen, O'Reilly, & Moore, 1994; Kirst, 1990). States and local districts have implemented outcomes-based accountability, which emphasizes the central role of student outcomes as a means of assessing school performance. The accountability systems typically incorporate multiple indicators and multiple forms of assessment; however, the core of these systems revolves around student performance. Outcome measures include students' test scores, self-evaluations, and teacher or peer evaluations.

What Is Accountability?

Hill and Bonan (1991) and Kirst (1990) have defined accountability as a systematic method to ensure that those working in a school system are moving in desired directions and thus are responsible to their constituents. Outcomes-based accountability means that the educational system is responsible for student attainment of specific learner outcomes. It differs from more traditional forms of accountability that have emphasized compliance with procedures or practices or have focused solely on input measures such as the resources allocated or expended for education (Brauen et al., 1994). Outcomes-based accountability requires more than assessing student educational outcomes or collecting other student data such as school participation and school completion rates. For schools to be held publicly responsible, student achievement, attendance data, and similar indicators of school performance must be reported and

measured against certain standards or criteria. In addition, accountability usually involves some explicit or implicit consequences for schools (e.g., principals or school staff) or individual students. For schools, these consequences can be in the form of rewards or sanctions that are administered based on how well students are meeting certain performance standards. For students, results of assessments are frequently linked to receipt of a high school diploma.

The premise of outcomes-based accountability is that school staff have to be guided by the cognitive, behavioral, and attitudinal results they produce in the students they teach.

> In sum, the premise of outcomes-based accountability is . . . that school staff have to . . . be guided by the cognitive, behavioral, and attitudinal results they produce in the students they teach. Outcomes-based accountability also means that outcomes not only must be measured, but they must have some consequences, (Brauen et al., 1994, p. 3)

Issues in Linking Student Outcomes to School Accountability

A number of issues must be addressed if schools are to be held accountable for student outcomes. These include questions of (a) how to define the valued outcomes, (b) how to assess the outcomes, and (c) what standards will be applied. It is also important to determine the consequences for individuals representing the school systems as well as the consequences for students.

These issues are posing major challenges to states and local school districts across the country. For the past two decades, individual states have been moving toward developing a systems approach to improving educational outcomes for students. By 1992, 46 states had developed some type of systemic approach to evaluating student outcomes (Council for Educational Development and Research, 1994). These endeavors have resulted in a variety of approaches to defining and assessing student outcomes. Some are very specific and relate to academic content areas, while others define learning or thinking processes or social or emotional behaviors (Warren, 1993). However, two common characteristics have pervaded virtually every outcome assessment system developed by the states: (1) an emphasis on increasing the expectations for student performance through the development of standards and

(2) increasing the consequences attached to that performance (Brauen et al., 1994; Warren, 1993).

Accountability for Outcomes of Students with Disabilities

The current general education activities related to outcomes-based accountability have major implications for students receiving special education services. Special educators are increasingly concerned with the outcomes of students who are identified as having disabilities and receiving special education support services (Hasazi, Gordon, & Roe, 1985; Jakweth & Frey, 1992; National Center on Educational Outcomes [NCEO], 1991; National Council on Disability, 1993; Wagner, 1991). These studies have examined the postschool status of varied groups of students with disabilities. The outcomes reported in these studies include higher than desired rates of unemployment, as well as low wages and limited community or social independence. These disappointing outcomes have prompted special education policymakers to call for increased attention to the results of special education programs.

Historically, special education program accountability has focused on determining whether a service has been provided as documented in the individualized education program (IEP) and on other procedural compliance indicators such as the timeliness of assessments and parental notifications. Whatever educational outcome data have been reported are incorporated into the IEP. However, research (Olsen & Massanari, 1991; Smith, 1990) indicates that the IEP has evolved into little more than a compliance document in many school districts. It documents services and placements but is of little use for instruction or evaluation. Moreover, the IEP was never designed to monitor progress toward long-term program outcomes.

The level of rigor applied to the assessment of outcomes for nondisabled students must also be applied to students who are receiving special education services.

As noted by DeStefano (1993), efforts to document the effectiveness of special education programs by assessing student attainment of long-term goals are almost totally nonexistent. DeStafano and others (McGrew, Thurlow, Shriner, & Spiegel, 1992; McLaughlin & Warren, 1994) believe that educational systems must be accountable for all students and that the level of rigor applied to the assessment of outcomes for nondisabled students must also be applied to students who are receiving special education services. Furthermore, there must be equally

rigorous standards for performance, and there must be consequences for the educational systems.

Lacking a system of comprehensive accountability, the educational system has no way of knowing whether or not a particular program is effective in educating students with disabilities. However, accountability is dependent upon having a system of assessments that is flexible, can accommodate diverse learners, and produces information that is valid and reliable (Jakweth & Frey, 1992).

2. Assessment of Student Outcomes

Increasingly, outcomes-based accountability systems are using results of both traditional assessments, such as norm-referenced tests, and authentic or performance assessments in an effort to measure broad domains of student knowledge.

Traditional Assessment

Norm-referenced tests have long been used to provide educators and policymakers with information on student progress. This form of testing is useful in ascertaining student skills across a large number of domains, in a relatively short period of time, and in a cost-effective manner. Furthermore, these tests are administered and evaluated under uniform conditions so that results can be compared across groups of students (Office of Technology Assessment, 1992).

Although multiple-choice and true/false questions are typically used to describe more traditional assessments, it is important to note that a variety of other testing formats may fall into this category. Examples could include oral examinations and written essays (Office of Technology Assessment, 1992).

Use of traditional assessments has been increasingly criticized in recent years because they lack integration of instruction and testing; emphasize factual knowledge and discrete skills; force the selection of one right answer versus exploration of multiple possibilities; require artificially short answers; and force students to work in isolation (Linn, Baker, & Dunbar, 1991; Office of Technology Assessment, 1992). In addition, traditional assessments have been challenged as being unfair to individuals of diverse cultures and learning styles (Rothman, 1991).

Moreover, these types of assessments are inconsistent with many of the current educational outcomes being defined by states. For example, student outcomes such as making judgments, solving problems, reasoning, or communicating for multiple purposes would be difficult to measure in traditional multiple-choice or short-answer formats.

Rather, assessment of a student's ability to demonstrate reasoning, gather information from multiple sources, and contribute to collaborative efforts requires a multifaceted approach.

In response to these concerns, a growing emphasis has been placed on developing student assessments that can measure knowledge and competencies in ways that require students to demonstrate the higher-order process skills as well as content knowledge.

Issues in Using Performance Assessments in Large-Scale On-Demand Assessment Programs

In response to the desire to measure student abilities in more complex and integrated learning environments, many states and local districts have turned to alternative assessments that provide a broader measure of student knowledge and skills. Of particular interest is the use of performance assessments that measure the extent to which students can integrate knowledge, apply it to meaningful and complex tasks, and produce a response, whether through demonstrations, constructions, or performances (Poteet, Choate, & Stewart, 1993).

Performance assessments may take a variety of forms, including essays, hands-on science problems, open-ended problems, computer simulations of real-world problems, artistic productions, and portfolios of student work. The assessments require the performance of an activity that integrates knowledge from a variety of domains (i.e., science, mathematics, reading, writing) and results in a product—either written, constructed, verbal, or performance (Office of Technology Assessment, 1992; Shavelson, Baxter, & Pine, 1992). Evaluations are based on the "right" answer as well as a defense for the rationale or process for developing the response (Shavelson et al., 1992).

Well-designed performance assessments can give teachers and students instructionally relevant information so the results can be used to adjust instruction (Office of Technology Assessment, 1992). Thus, policymakers view the use of these tests as a way to directly influence curriculum and teacher behavior. The reasoning is that if schools are held accountable for students' proficiencies, then teachers will teach those critical skills and behaviors.

A push for a national commitment to performance assessment is being spearheaded by FairTest: The National Center for Fair and Open Testing (Rothman, 1991). Representatives of over two dozen education, civil rights, and advocacy groups were called on by FairTest to work toward the development of a national assessment system focusing on reliable performance assessments that are sensitive to the cultural biases found in existing assessments (Rothman, 1991). The ultimate goal of this

group is to ensure that national assessments directly address and are sensitive to cultural biases.

When used for accountability purposes, the assessments must be conducted with a large number of students, there must be consistency in the domains of knowledge being assessed, and the assessments must yield adequate samples of student performance within those domains.

However, there are a number of challenges when performance assessments are used as part of large-scale federal, state, or district-wide assessment programs. When used for accountability purposes, the assessments must be conducted with a large number of students, there must be consistency in the domains of knowledge being assessed, and the assessments must yield adequate samples of student performance within those domains. Costs and time associated with administration of the assessments to large numbers of students are also major considerations. Finally, reliability in scoring the assessments becomes critical, especially if the results are linked to sanctions or rewards for schools or consequences for students, such as the award of a diploma. Two issues that are particularly troublesome when considering large-scale use of performance assessments are costs and scoring.

Costs

Performance assessments and other forms of authentic assessments have one significant drawback when used as part of large-scale on-demand assessment programs: their cost. Authentic assessments are costly to develop as well as to administer and score. While some of the expense lies with equipment and supplies to be used in the performance tasks, the majority of the costs are incurred in personnel. Significant investments of human capital are required, particularly in the areas of task administration and scoring of the tasks (Office of Technology Assessment, 1992; Popham, 1993).

One approach to maximizing the balance among time, cost, and desire to obtain wide-scale data on the progress of students throughout a program or system is the concept of *matrix sampling* (Popham, 1993). This approach enables assessment and accountability designers to systematically sample an array of students and performance tasks. No student performs all tasks, yet there are sufficient numbers of performance tasks over all to enable the system to evaluate programs. Such sampling reduces costs and time burdens but limits the ability to link

scores to individual students, since not all students are assessed. This limits the use of the assessments for improving instruction of individual students.

Scoring the assessments is also costly. Because student performance is not reduced to a correct answer on a test form, a number of individuals must be trained and employed as assessors. This, in turn, requires investment of resources to train the assessors.

Reliability in scoring performance assessments is a major issue, particularly when these forms of assessment are used in large-scale programs.

Scoring

Many of the issues surrounding the scoring of performance assessments have been discussed in other books in this Mini-Library. Reliability and validity are two critical aspects of effective assessments. Reliability in scoring performance assessments is a major issue, particularly when these forms of assessment are used in large-scale programs. Because teachers or other trained assessors are asked to rate diverse types of student work, intensive training is required. However, some (Baker & Linn, 1993; Dietel, 1993) believe that current scoring techniques may not be strong enough to warrant use of some performance assessments for school-wide accountability. As evaluators search for innovative ways to assess students' abilities to integrate knowledge and skills in meaningful ways and to demonstrate higher-order thinking skills, it becomes more challenging to ensure that assessments are reliable and valid.

3. Issues in Using Performance Assessments for Students with Disabilities

Performance assessments can be extremely relevant in the students' lives while offering educators the opportunity to evaluate student progress and program effectiveness.

Performance assessments offer a number of benefits over the use of traditional standardized assessments for students with disabilities (Coutinho & Malouf, 1993; Poteet, Choate, & Stewart, 1993). Perhaps most important is the potential for linking instruction and assessment. Since authentic performance assessments are based on performance of valued skills, the experience of completing a task may be intertwined with the instructional process. Thus, performance assessments can be extremely relevant in the students' lives while offering educators the opportunity to evaluate student progress and program effectiveness.

However, use of performance assessments as part of a system's accountability plan raises a number of issues for students with disabilities. These include (a) defining the outcomes to be assessed, (b) developing performance standards, (c) developing assessment accommodations, and (d) scoring.

Defining the Outcomes

Assessment programs are constructed to measure progress toward valued educational outcomes. When outcome frameworks are defined too narrowly (e.g., academic content domains) and neglect other valued areas (e.g., vocational skills, personal management, social skills, and communication), the outcomes may not reflect all of the skills and competencies that are considered appropriate for students with disabilities. In particular, students with significant cognitive disabilities may

have great difficulty demonstrating knowledge in academic content areas as the curriculum for these students has focused on developing functional life skills. As a result, most students with moderate to significant cognitive disabilities have not had access to the instruction that prepares them for successful completion of the assessment tasks. Moreover, the tasks are not relevant to the students' program goals.

Developing Performance Standards

Standards are the benchmarks against which student performance may be compared. A critical decision in designing an assessment system is whether students will be compared to themselves to determine change in their performance over time, or whether they will be compared to fixed standards of performance (Brauen et al., 1994).

A critical decision in designing an assessment system is whether students will be compared to themselves to determine change in their performance over time, or whether they will be compared to fixed standards of performance.

Fixed standards promote high expectations for all students and set criteria that can be used consistently across scoring rubrics. Many students with disabilities cannot meet some of the absolute standards, particularly in academic skill areas. Where such standards exist, students with disabilities may be excluded or discouraged from participating in the accountability system. When participation in the assessment program is linked to high school diplomas, students with disabilities may be at a particular disadvantage.

Assessment Accommodations

Many students with disabilities will require accommodations to participate in large-scale assessment programs. These accommodations may include additional time to complete the task, alternative testing locations, alternative means of administration (e.g., reading, interpretation, braille), and alternative supplies or equipment (e.g., computers). Some students, particularly those with significant cognitive disabilities, may require alternative forms of assessments. While the provision of these accommodations will result in a more inclusive assessment system for all students, there is concern that the accommodations will be costly and,

more important, that they will compromise the integrity of the assessment and jeopardize the reliability of the results.

Scoring

Related to assessment accommodation is the issue of scoring. When assessment results must be reported in the aggregate and when results matter, scoring reliability becomes more critical. Assessment experts are reluctant to permit accommodations because they believe they compromise the assessments. Rigorous scoring is as necessary in alternative assessments as in traditional assessments. However, the nature of student products and performances for students with moderate to significant disabilities who participate in these alternative programs makes it extremely difficult to develop scoring rubrics that allow assessments to be scored reliably.

4. Performance Assessments in Action

Performance assessments are being incorporated into accountability systems with increasing frequency. A number of state and local school districts have adopted this approach as either the sole method of assessment or in conjunction with traditional norm-referenced tests. Many states and local school districts have included students with disabilities in their accountability systems (see Thurlow, 1994). In an effort to understand the implications of including students with disabilities in these assessment programs, telephone interviews were conducted with representatives of three states (Kentucky, Maryland, and Vermont) and two local school districts (Littleton, Colorado, and Arlington Heights, Illinois). These jurisdictions were selected because of the extent of their commitment to using performance assessments for statewide or district accountability.

Interviews were conducted with key informants in each of the states or local districts to obtain information relative to the general use of performance assessments, as well as specific issues related to students with disabilities. In addition, documents describing the systems were reviewed. Following are brief descriptions of the five assessment programs.

Kentucky

Description of Accountability System

The Kentucky Educational Reform Act (KERA) outlines six performance goals that all students are expected to attain upon graduation from Kentucky schools (see Table 1). Individual schools have been granted significant latitude in determining how students will master the 75 state-developed student outcomes. These outcomes apply to all students, regardless of the curriculum they may be receiving (special education, vocational, or academic), and they include academic as well as personal management and social skills.

TABLE 1
Characteristics of Selected Accountability Systems Using Performance Assessments

State	*Outcomes*	*Assessments*	*How Results Are Used*	*Inclusion of Students with Disabilities*
Kentucky	Six performance goals include communication and math, core concepts from the sciences, arts, humanities, social studies, and practical living studies; self-sufficiency; membership in family, work group, or community; thinking and problem solving; and connecting and integrating knowledge. Seventy-five outcomes in the areas of reading, writing, mathematics, science, social studies, arts and humanities, and vocational education/practical living.	Portfolios in writing and math for all students in grades 4, 8, and 12; Performance events for all students in grades 4, 8, and 12. Focus on mathematics, science, social studies, arts and humanities, and vocational education/practical living; Transitional assessments (open-ended and multiple-choice questions) for all students in grades 4, 8, and 12. Focus in reading, writing, mathematics, science, social studies, arts and humanities, and vocational education/practical living; or Alternative Portfolios for students with severe disabilities in grades 4, 8, and 12. Maximum of 2% of the school enrollment may develop Alternative Portfolios.	School report cards. School rewards and sanctions.	All students are required to participate in the transitional assessments or Alternative Portfolios unless a physician provides a statement documenting significant negative impact on the student's health as a result of participating.

continues

TABLE 1 (continued)

State	Outcomes	Assessments	How Results Are Used	Inclusion of Students with Disabilities
Maryland	Content outcomes of reading and writing, mathematics, social studies, and science. Overarching outcomes of individual and group problem solving, decision making, and issue analysis.	Maryland State Performance Assessment Program (MSPAP) (matrix sampling selection of students in grades 3, 5, 8). Areas include reading, mathematics, writing and language usage, social studies, and science; Comprehensive Test of Basic Skills (CTBS) (matrix sampling of students in grades 3, 5, 8); Maryland Functional Tests (to be passed by graduation). Areas include reading, mathematics, writing, and citizenship; or Proposed: Alternative Performance Assessments for students with significant cognitive disabilities ages 9, 11, 14, 17, and 21.	Annual school performance reports. Annual school district performance reports. Annual state performance reports. Student diplomas. School rewards and sanctions.	Students with disabilities may be exempted by their IEP team from participation in MSPAP, CTBS, and Functional Tests. Students with significant cognitive disabilities are eligible to participate in the Alternative Performance Assessments.
Vermont	Fields of knowledge outcomes include mathematics; life science; physical science; earth science; technology education; history and philosophy of science; and career preparation in sciences, math, and technology. Vital results outcomes include communication, reasoning and problem solving, personal development, and social responsibility.	Uniform Assessment (math) includes two 40-item multiple-choice tests and a single on-demand task for students in grades 4 and 8; Uniform Assessment (writing) requires students (grades 4 and 8) to respond to a single prompt in a 90-minute period; and Math and writing portfolios for students in grades 4 and 8.	Supervisory union reports. State reports.	School-based staffing teams determine whether or not students with disabilities are eligible for exemption from the portfolio development and assessments. Exempted students may build portfolios that do not conform to the state requirements for use in instruction.

continues

TABLE 1 (continued)

State	*Outcomes*	*Assessments*	*How Results Are Used*	*Inclusion of Students with Disabilities*
Littleton, CO*	Seven system outcomes that include the ability to gather, process, and communicate information through reading, written and oral expression, and active listening; mathematical concepts and processes; cultural, scientific, and technological literacy; critical thinking, problem-solving and decision-making skills, and intellectual curiosity; creative expression; physical and emotional well-being; and self-esteem and respect for others.	Individual schools or school clusters are permitted to develop their own performance assessments, which must link to the seven system outcomes.	School report cards. Student diplomas.	Students with disabilities who spend 50% or more of the day in the regular classroom are expected to participate in the assessment program.
Arlington Heights, IL	Eleven General Learner Outcomes (GLOs) include communication, social interaction, analysis, problem solving, value judgments and decisions, creativity, civic responsibility, global environment, wellness, technology, and life and career planning.	Standardized achievement tests (Iowa Test of Basic Skills in grades 2, 5, and 8; and Tests of Achievement and Proficiency in grades 9 and 11); Standardized ability measures (Cognitive Abilities Test in grades 2, 5, 8, 9, and 11); Integrated Reading and Writing performance assessments (grades 4 and 8); Writing assessment (grade 12); and A variety of school-created performance assessments administered in grades K–12.	Individual student matrices. Student diplomas.	Students with disabilities may be exempted from the traditional assessments. However, since the performance assessments are school created, all students are expected to participate in assessments that have been developed based on their individual skills and needs.

*Program dismantled during 1993–1994 school year due to political pressure.

Annual school report cards indicate progress toward meeting the state outcomes by listing the percentage of students scoring at each of four performance levels (novice, apprentice, proficient, distinguished) in each of four subject areas (reading, math, science, and social studies).

School districts are given the flexibility to develop instructional content and approaches appropriate for their students. The Kentucky Instructional Results Information System (KIRIS) develops an array of assessments to determine student progress. Currently, progress is evaluated based on performances in the KIRIS assessments or the Alternate Portfolio Project, a multidisciplinary approach to evaluating student progress in meeting the valued outcomes.

The local school is held accountable. Annual school report cards indicate progress toward meeting the state outcomes by listing the percentage of students scoring at each of four performance levels (novice, apprentice, proficient, distinguished) in each of four subject areas (reading, math, science, and social studies).

Assessments

KIRIS contains three types of assessments: (1) portfolios and exhibitions, (2) performance-based events in which students apply what they have learned to a real situation, and (3) transitional assessments including multiple-choice and open-ended questions. Participation in the state-developed mandated components of KIRIS are administered in grades 4, 8, and 12. The state is also developing assessments for use on a voluntary basis in all grades.

For the 1991–1992 school year, Kentucky students participated in the writing portfolio process; transitional assessments (i.e., multiple-choice, open-response testing in mathematics, social studies, science, and reading); and performance events testing mathematics, science, and social studies. During the 1992–1993 year, students also developed mathematics portfolios. Plans for subsequent school years were to focus on the development of integrated, holistic portfolios.

The vast majority of students with disabilities participate in this mainstream assessment system. However, a small percentage (1%–2%) of students with disabilities are selected to participate in the Alternate Portfolio Project. These students generally have moderate to significant cognitive disabilities and are not working toward a high school diploma. Students are selected to participate entirely in one system or the other.

Since the 1992–1993 school year, students with severe disabilities have participated in the Alternate Portfolio Project. Their entries include a written or dictated letter of introduction describing the portfolio; relevant academic work; schedules reflecting time with nondisabled peers; résumé of job experiences; samples of their present mode(s) of communication; and letters of validation from educators, family members, and other supporters. Twenty-eight outcomes were selected as preliminary targeted outcomes for the students participating in the Alternate Portfolio Project. Additional outcomes will be added each year until all students' portfolios include work reflecting progress toward all 75 valued outcomes.

Exemptions

All students, including those receiving instruction in a home or hospital setting, are required to participate in the KIRIS or Alternate Portfolio Project unless a physician provides a statement documenting significant negative impact on the student's health as a result of participating. These students are exempted from participation in the assessment and accountability aspects of KERA.

Accommodations

A wide array of accommodations is permitted for students with disabilities. The state education agency (SEA) has provided significant latitude in identifying acceptable accommodations, the critical issue being that the IEP team must approve the accommodations and they must be used by the student on a daily basis in the course of ongoing instructional activities.

The primary focus during portfolio evaluations is the student's ability to perform a series of tasks that result in a routine that is meaningful in the community. Therefore, less focus is placed on supports used by a student (e.g., calculators, dictionaries, electronic spell checkers) than would be found in more traditional assessments. Accommodations provided by nondisabled peers (e.g., student peers, coworkers, store workers) are strongly encouraged.

General Comments

Educators and advocates in Kentucky have learned that the new approaches to assessing student performance have created a need for new ways of teaching. This has resulted in a need for teacher training. Initially, training focused on discrete topics related to the curriculum in 1-hour workshops, but teachers are now beginning to view these ses-

sions not only as vehicles to acquire new knowledge but also as opportunities to reflect on instruction and share ideas with colleagues.

Experiences with KIRIS activities have promoted some major changes in the education of students with severe disabilities. Since inclusion of students with disabilities in school and the community is favored in the scoring rubrics, educators have been strongly encouraged to become more community based in their instructional practices. While community-based instruction and inclusive education started in response to the assessments, many educators are seeing the inherent value of these for students with disabilities and those with diverse learning styles.

Maryland

A School Performance Review System serves as the accreditation body to reward schools for outstanding performance, recognize those that are progressing, and implement sanctions for those that are not meeting standards.

Description of Accountability System

The Maryland School Performance Program (MSPP) was developed as a comprehensive student outcomes accountability system. Reflecting state-level goals and strategies (see Table 1), student learning outcomes have been developed in the areas of reading and writing, mathematics, social studies, and science.

MSPP communicates results to the public through school report cards (Annual School Performance Reports) that provide data related to student population characteristics, student participation rates, and the results of selected student assessments. School reports are aggregated for each local school district. Outcomes data for students with disabilities participating in MSPP assessments are aggregated with those of their nondisabled peers.

A School Performance Review System serves as the accreditation body to reward schools for outstanding performance, recognize those that are progressing, and implement sanctions for those that are not meeting standards. Progress benchmarks have been established for individual schools with the expectation that all schools will reach state standards by 1995, and sanctions—including takeovers by local school

districts (i.e., reconstitution)—will be imposed on schools that do not meet the standards.

Assessments

Assessment of student outcomes within the MSPP includes both norm-referenced tests (i.e., Comprehensive Test of Basic Skills [CTBS]), criterion-referenced tests (i.e., Maryland School Performance Assessment Program [MSPAP]), and the Maryland Functional Tests.

The CTBS is administered to a matrix sampling of 3rd, 5th, and 8th grade students.The Maryland Functional Testing Program is a series of four criterion-referenced tests designed to ensure that students possess basic competencies and functional skills in mathematics, reading, writing, and citizenship. Students must pass each of the four tests as a requirement for high school graduation. With the exception of the Maryland Writing Test, multiple-choice items are the primary response mode.

The MSPAP is a collection of performance assessments created to provide a mechanism for measuring the desired student outcomes identified within MSPP. Performance tasks are administered to a matrix sampling of students in grades 3, 5, and 8, unless they are exempted from participation (i.e., due to disability or limited English proficiency). The tasks are designed to measure what students have learned based on the identified state outcomes. The assessments are integrated across subject matter content and emphasize the thoughtful application of knowledge and skills in content areas. For example, a task for 3rd graders assesses student skills in social studies and language usage through the analysis of a scenario of welcoming a new student from Japan. Students read excerpts on Japanese schools, and are asked to construct maps of the new school, develop a schedule that includes comparisons to the Japanese schedule, and write an analysis of the adjustments the student will face in the new school. Each assessment requires multiple student responses that can include open-ended responses, writing mathematical findings, and group activities. Results are not reported for individual students. Rather, they are used to provide descriptive data about a school's performance in the assessed areas.

Exemptions

Although the outcomes developed by the state are sufficiently broad to include most students, MSPAP assessments reflect a focus on higher-order, critical thinking skills within specific academic areas. As a result, they do not match the curriculum goals of some students. Therefore, exemptions may be granted for non-diploma-seeking students with disabilities and (on a one-time basis) students with limited English

proficiency. Exemptions for students with disabilities are made on a student-by-student basis through the IEP process.

An Alternative Performance Assessment system is being developed by the state for students with significant cognitive disabilities who are exempted from components of MSPP. The core of these assessments will be portfolios, but they will include on-demand performance assessments that will evaluate student performance and program supports in four domains: vocational, recreation/leisure, personal management, and community. It is anticipated that these assessments will become a part of MSPP and the Annual School Performance Reports for each local district, thereby ensuring accountability of special education programs and services for students with significant cognitive disabilities.

Accommodations

The Maryland State Department of Education, in collaboration with a network of special education teachers, has developed a list of permissible accommodations for each of the MSPP assessments. Accommodations (e.g., scheduling, setting, equipment, presentation, and response) are permitted only when they do not invalidate the assessment for which they are granted. Accommodations are to be based on individual student needs rather than category of disability, level of instruction, or other group characteristics. The accommodations are developed during the annual IEP team meeting and are to be used throughout the course of ongoing instruction during the school year.

Placing accountability for student achievement directly at the school level is creating a new sense of ownership for students and a greater awareness among school staff of the unique skills and needs of many students and how these affect school performance results.

General Comments

While the Annual School Performance Reports are key to school restructuring, the MSPAP is viewed as the vehicle for effecting change in instructional strategies to improve student outcomes. Teachers are talking of new educational values (e.g., critical thinking skills and integrated thinking) and new ways of teaching and testing students (e.g., performance-based assessment and development of meaningful products).

MSPP is having a significant impact on the process of school-site restructuring throughout the entire state. Placing accountability for student achievement directly at the school level is creating a new sense of

ownership for students and a greater awareness among school staff of the unique skills and needs of many students and how these affect school performance results.

Vermont

Description of Accountability System

The Vermont Assessment Program was implemented statewide during the 1991–1992 school year. Vermont was the first state to adopt portfolios as part of a statewide assessment program. The program employs both standardized assessments and portfolios to collect information about the performance of 4th and 8th graders in mathematics and writing (see Table 1). Every student is given an opportunity to build a portfolio. The portfolio scoring establishes clear performance standards, and portfolios are scored by trained teachers using standard rubrics. Portfolios are evaluated according to several criteria on a four-point scale. They include classroom work completed throughout the year, as well as a "best piece" chosen by the student.

Assessments

Information about the statewide performance of students in mathematics is gathered from the Uniform Assessment and portfolios. The Uniform Assessment includes two 40-item multiple-choice tests and a single on-demand task that the student must complete independently. Each student's portfolio is expected to include 10 to 20 items. From these, students select 5 to 7 "best pieces" to be scored. The mathematics scoring system evaluates student responses using four criteria related to problem solving and three related to mathematical communication. For each criterion, each of the four performance levels is identified with a benchmark or student paper that serves as an example of a performance rating of 1, 2, 3, or 4.

The writing assessment contains information from three sources:

1. The Uniform Assessment requires all students in a grade level to respond to a single writing prompt within 90 minutes.
2. The "best piece" is selected by the student from the portfolio contents to represent his or her best work as a writer.
3. The remaining contents of the portfolio are scored together and referred to as the "portfolio."

All student writing is evaluated on the basis of criteria on five dimensions: purpose, organization, detail, voice/tone, and gram-

mar/usage/mechanics. The writing assessment describes levels of performance in terms of the frequency with which students display the skills: rarely, sometimes, frequently, or extensively. For each criterion dimension, student papers have been selected to serve as benchmarks or illustrations of each performance level.

Scores for portfolios are reported at the supervisory district level. In the future, scores for the uniform writing piece will be reported school by school. However, the results of the portfolio assessments are clearly tied to instruction. Scoring standards define for teachers, students, and parents what is expected in the two subject areas, and the standards are being used to modify individual students' programs.

Exemptions

The SEA wants all students to participate in the portfolio assessment, but the basic staffing team in each school can decide whether participation would be harmful or the portfolio content is not relevant for a particular student. Reportedly very few students are exempted, but each district handles this differently. Students with disabilities who are exempted may build portfolios that do not conform to the state requirements. These are scored by the student's teacher and are used only for instruction. Local school personnel report that this process occurs infrequently.

Accommodations

Accommodations are permitted for any student as long as the specific accommodation is used as part of instruction. For example, if a student uses a computer for writing instruction, then the student may use the computer for any part of the writing tasks.

Teachers reported that portfolios had both positive and negative effects on the performance of students of low ability.

General Comments

There is reportedly little resistance among teachers and administrators to the use of the portfolio assessments, because many schools had been using them for some time for instructional improvement in grades K through 12. The assessments are perceived as ways of improving individual student educational programs, not as ways of evaluating schools. There is also little reported resistance to the notion of accountability.

Vermont's assessment program is being thoroughly evaluated under a contract with the RAND Corporation (Koretz, Stecher, & Deibert, 1992). During the 1991–1992 pilot year, evaluators administered questionnaires to participating teachers and mathematics portfolio raters. Additional teacher questionnaires were administered and interviews were conducted with principals during the first year of statewide implementation (1991–1992). Interim analyses of these data indicated that portfolio raters had mixed reactions to the portfolio scoring process. Portfolio raters for both 4th grade and 8th grade were concerned about the unevenness of tasks that were submitted; they either were difficult to fit into a scoring category or were insufficient. Portfolio raters for 4th grade also had a difficult time actually scoring the tasks. Principals viewed the portfolio process as worthwhile but somewhat burdensome.

Teachers reported that they needed more training and clearer expectations about how to construct portfolios. They also indicated that they spent an average of 6 hours per week working on portfolios; half of the time was spent preparing portfolios and the other half was devoted to classroom portfolio activities. While time was a problem, teachers considered the effects on instruction to be great. Finally, teachers reported that portfolios had both positive and negative effects on the performance of students of low ability. Approximately 16% of the teachers reported that these students were often more successful as a result of portfolios. However, the majority of teachers indicated that the students had difficulties with the tasks.

Reliability of the portfolio scoring was also examined for the school years 1991–1992 (Koretz, Stecher, & Deibert, 1992) and 1992–1993 (Koretz, Klein, McCaffrey, & Stecher, 1993). Spearman correlation coefficients were computed between scores of two raters. Scores for individual pieces, as well as dimension-level (e.g., problem-solving) scores, were computed. For 1992–1993, the correlations for total scores for mathematics portfolios were .72 for grade 4 and .79 for grade 8. Similarly, correlations for writing total scores were .56 for grade 4 and .63 for grade 8.

Based on information from this comprehensive and ongoing evaluation, the SEA is moving to ensure greater reliability in scoring and to continue to provide intensive staff development to teachers.

Littleton, Colorado

Demographic Background

Littleton is a suburb of Denver. There are 16,500 students in the school system, of whom 1,500 are identified as needing special education. The school population is primarily white and middle class. The dropout rate

is around 4% for the system, and 88% of Littleton's graduates go on to college.

Description of Accountability System

Littleton's outcomes-based accountability system evolved from a strategic planning process that began during the 1980s. Seven broad educational outcomes were identified (see Table 1). These outcomes are assessed using both norm-referenced tests and performance assessments.

Individual school report cards are required, but schools have some latitude in what they report. All school report cards contain results of the norm-referenced tests. Performance assessment results are reported differently. For example, one school report card may contain a narrative description of its program. Diplomas are granted on the basis of success in attainment of outcomes.

Assessments

Individual schools or school clusters are allowed to identify their own outcomes as well as their own performance assessments, which must link to the seven system outcomes. As an example, Littleton High School, one of three district high schools, no longer uses Carnegie Units or required coursework as part of the graduation requirements. The school staff has established 19 outcomes and constructed 39 performance tasks, as well as a set of "exit" performance assessments, that are required to obtain the diploma. Throughout their 4 years of high school, students maintain a portfolio that contains the tasks as well as other work. The final task is a portfolio review conducted by a panel of school faculty during which the student presents and discusses his or her portfolio.

The other high schools are developing similar programs, resulting in three different sets of graduation requirements in the district. Middle and elementary schools are also developing outcomes and performance assessments, either individually or as part of a cluster of schools.

District-wide performance assessments include samples of student-selected nongraded or corrected work from 5th, 8th, and 11th grade classrooms. There are no formal rewards or sanctions for schools attached to the student performance data. The district has, however, supported principals for their involvement in the program through providing professional opportunities such as speaking at conferences.

Exemptions

Students identified as having disabilities may or may not take part in the district's total assessment program. If a student spends 50% or more of

his or her day in the regular classroom, the student is expected to participate in the assessment program. In addition, school staff must determine that a group test (e. g., the Iowa Test of Basic Skills) would be an appropriate measure of a student's progress. All but an estimated 1% to 2% of special education students participate in the assessments. For students who are exempted, the IEP is the accountability document.

The IEP team decides which school outcomes will be appropriate for the student and which of the assessments the student will take. However, no student may receive a diploma without successfully completing all of the exit performance assessments.

The district's IEPs are developed using the seven district outcomes. The IEP team decides which school outcomes will be appropriate for the student and which of the assessments the student will take. However, no student may receive a diploma without successfully completing all of the exit performance assessments.

Accommodations

Some assessment accommodations, such as braille assessments and allowing extra time, are permitted. In addition, some IEP teams have looked at whether the standards should be modified; however, assessment staff are critical of modifying the standards or providing testing accommodations.

There has reportedly been a great deal of discussion among special and general educators regarding students with disabilities and a recognition that most of these students can share the same outcomes as non-disabled peers.

General Comments

Recently there was a political backlash in the district, primarily centered around awarding the diploma on the basis of assessments that are considered to have insufficient validity and reliability. In addition, some parents were concerned about linking the "process" outcomes to the high school diploma. In the 1993 school board election, almost all proponents of the program were defeated. The new school board was

elected in the fall of 1993, and the district has decided to move away from outcomes-based education. The superintendent has retired, and a more traditional approach to granting diplomas is being reinstituted. However, individual schools are permitted and encouraged to continue to use performance assessments for instructional improvement.

Special education teachers were not initially interested in getting involved in the process of developing assessments. The teachers viewed the activity as an additional paperwork burden. Furthermore, special education teachers had not been involved in identifying outcomes or developing initial assessments so they believed there would be little relevance for students with disabilities. However, as those teachers have become involved, there has reportedly been a great deal of discussion among special and general educators regarding students with disabilities and a recognition that most of these students can share the same outcomes as non-disabled peers. There has also been some critical scrutiny of the teaching and learning process in special education, and those teachers report that a lot of extraneous instruction that was provided in special education classrooms has been eliminated. Teacher involvement and collaboration were cited as among the most positive aspects of the entire program.

Arlington Heights, Illinois

Demographic Background

Arlington Heights District 214 is a grade 9 through 12 district with six high schools. Arlington Heights is a suburb of Chicago with 11,000 students, about 1,000 of whom are identified as having disabilities. The student population is economically and culturally diverse. In one high school over 50 languages are spoken; another school's majority population is Hispanic; and one high school is over one third Jewish.

Description of Accountability System

The district embarked on establishing an outcomes-based education system in 1990. There was wide stakeholder input—including specific attention to community clergy and other key groups—concerning district plans. There are two parts to the district's assessment plan. The first involves 11 general learner outcomes (GLOs) (see Table 1). An interdisciplinary team of teachers and administrators developed a set of indicators for each of the GLOs. Students must demonstrate the following three levels of achievement for each outcome:

Level 1: The knowledge base related to the outcomes.

Level 2: Practical knowledge application with various contexts.

Level 3: Ability to transfer learning to new situations.

In addition, students are required to demonstrate competency in each outcome in a minimum of three different content areas. Beginning in 1995, students will be awarded high school diplomas based on their attainment of the GLOs. The second component of the system includes state-mandated assessments in six program areas.

Assessments

Students demonstrate competency through performance assessments that are created and evaluated by individual teachers. There are no uniform tasks and no permanent records of student performance, other than those that a teacher may choose to keep. Student progress toward outcomes is monitored on a matrix of the GLOs and the levels of knowledge. Teachers may use portfolios, performances, and teacher-made tests. There is also a bank of teacher-made tests at the district level from which teachers may choose. Principals and department heads informally monitor classroom teachers to ensure that outcomes are being assessed, but there are no specific standards. The district is developing a generic rubric that might be applied to the individual performance assessments.

A second part of the district's assessment is required by the State of Illinois. Each student must be assessed in six program areas (English, math, social science, physical education, health, fine arts). For these outcomes, District 214 is developing performance tasks and criterion-referenced assessments that will be used consistently across the district.

Exemptions

The district has resource room programs for students with mild disabilities, one self-contained classroom for students with behavior disorders, one program for deaf students, and an alternative high school. Students with severe disabilities are educated in a regional program and are not included in the assessments. The SEA mandates that all students be assessed in the six goal areas, so the district is defining goals within each area that are appropriate for diverse learners. For example, English now includes speaking and communication and listening and looking at books in addition to reading and writing. Students with significant cognitive disabilities need only show progress toward the state goals, not mastery.

For the district's GLOs, all students must demonstrate attainment; however, the assessments are tailored to individual students. In addition, the interpretation of an outcome or indicator is also individualized based on input from the special and general education teachers.

Accommodations

Any necessary accommodations are included in the assessments designed for a particular student. Since there are currently no uniform performance tasks, specific accommodations are not an issue.

The district's approach is designed to be nonthreatening and to focus on providing students with many opportunities to demonstrate outcomes, thereby enriching the curriculum across content areas.

General Comments

The district has opted for a flexible assessment system in large part to obtain critical stakeholder endorsement. Standards, administrative burdens, record keeping and other paperwork, and school accountability are still minimal, and school staff are being encouraged to experiment and innovate. The district's approach is designed to be nonthreatening and to focus on providing students with many opportunities to demonstrate outcomes, thereby enriching the curriculum across content areas. Teachers have opportunities for professional development, including weekly meetings to share assessments and discuss issues in administering performance assessments.

5. Challenges and Implications: Using Performance Assessments for Accountability

General Issues

The use of performance assessments as part of accountability systems poses a number of challenges. In general, the issues of program costs and reliability in scoring need to be considered when using performance tasks as part of large-scale assessment programs. According to those interviewed, the areas that appeared most difficult over all were setting performance standards and establishing scoring rubrics that are fair, yet flexible. These issues are potentially more important when the results of the assessments are used to make decisions about high school diplomas or to sanction schools. An additional issue was deciding what to assess. Specifying the outcomes and the indicators was a difficult endeavor, both politically and conceptually.

In all sites, the total assessment program included the use of more than performance assessment tasks. In some instances, the use of norm-referenced assessments was mandated by state policy. In other sites, traditional norm-referenced assessments were used because some educators and policymakers believed they provided a greater sense of security in gauging student progress. These assessment results also permitted cross-district or cross-site comparisons. However, performance assessments were considered the core of the assessment programs.

In Maryland, the MSPAP performance tasks were administered consistently across the state to a sample of students. In other jurisdictions, performance assessments included some consistently administered tasks along with other samples of student work (e.g., portfolio entries). These could vary from school to school. In all sites, there was a major emphasis on making certain that the assessments provided many opportunities for students to demonstrate proficiency in an outcome area and that results could influence instruction either individually or at the school level.

While those interviewed did not mention authenticity as a major reason for using performance assessments, all did acknowledge that educational outcomes considered important in their states or districts go well beyond knowledge of traditional subject matter and include the application and integration of subject matter content as well as problem solving, making judgments, and communicating. These educational outcomes could not be measured by more traditional assessments; they required the use of performance tasks. In addition, all those interviewed spoke of how performance assessments have benefited student learning because the assessments directly link to curriculum and can change classroom instruction. That feature, more than any other, was prompting policymakers to maintain performance assessments as the core of the assessment program.

Teachers, in particular, were cited as benefiting from the use of performance assessments in several ways. First, in every site, classroom teachers are being involved in some level of designing or implementing the assessment program. In some districts, teachers are helping to define the outcomes and develop and score assessments. Teachers are working collaboratively to develop curriculum-embedded performance assessments. This process is promoting teacher collaboration and fostering self-directed professional development. Special educators have become involved in this collaboration and have been challenged to reevaluate their expectations for students with disabilities, as well as the content of their instruction.

A difficult conceptual switch for some special educators is the move from believing that each student with a disability should have individualized outcomes to accepting the notion of a common set of outcomes across students.

Students with Disabilities

All of the districts and states studied (with the exception of Kentucky) permit students with disabilities to be exempted from participation in the assessment programs. Typically, decisions to exempt these students are made by the individual IEP teams. Exemptions most often are made if the team perceives that the assessments would be particularly frustrating or cannot be performed by the student or if the outcomes being assessed are not relevant to the student's educational program. Despite the strong impetus to include all students with disabilities in the assessments, there is still some ambiguity regarding how much individualization should be provided within the assessment program. A difficult

conceptual switch for some special educators is the move from believing that each student with a disability should have individualized outcomes to accepting the notion of a common set of outcomes across students. In some instances, this means accepting the notion that students with disabilities should be permitted to have opportunities to attain the same valued educational outcomes expected of nondisabled students and to experience a broad and balanced curriculum. In the case of students with significant cognitive disabilities, separate or additional outcomes may need to be established that can set the broad program goals for all of these students.

There is a risk that many students experiencing learning problems, including those identified as having disabilities, may not receive high school diplomas if they must meet certain set performance standards on the assessments.

Defining scoring standards for students with disabilities is also problematic in most sites included in the study. In Maryland, Vermont, Kentucky, and Littleton, Colorado, which have standard scoring systems, assessment accommodations are permitted for students with disabilities, but these students are held to the same performance criteria as nondisabled students. Individual standard setting occurs in Arlington Heights, Illinois. In some systems, accommodations include administering alternative assessments and even defining alternative outcomes. These types of accommodations are made to ensure participation of students with significant disabilities in the assessment program and presumably to inform instruction. The results of these assessments are not generally used for program accountability or to grant diplomas. Linking diplomas to student assessment results has already been challenged for all students in Littleton, Colorado, and it is too early to know what the effects will be in Arlington Heights, Illinois. Nonetheless, there is a risk that many students experiencing learning problems, including those identified as having disabilities, may not receive high school diplomas if they must meet certain set performance standards on the assessments.

Setting performance standards for students w th disabilities is an unresolved issue, and it frequently results in the exclusion of students with disabilities from high-stakes assessments (NCEO, 1991). While DeStefano (1993) has cited a number of reasons why students with disabilities should be included in assessment and accountability initiatives, she has acknowledged the challenges. These include identifying meaningful outcomes, defining performance standards in sufficiently

broad terms, and creating enough flexibility in the assessment system to accommodate the unique needs of students with disabilities. Others (e.g., Jakweth & Frey, 1992; Richards, 1988; Weber & Zin, 1992) also have cited the need to have a flexible assessment system that employs multiple data-gathering strategies including on-demand assessments, examples of student work, and teacher judgments.

It appears from the experiences of the five states and local districts surveyed that the use of performance assessments as part of a larger accountability effort has permitted more students with disabilities to be included in the assessment program. Equally important, inclusion in the assessment program means increased attention to outcomes and increased scrutiny of the instruction provided to students with disabilities. Furthermore, the assessments are leading special educators to reconsider some of their expectations for students with disabilities and to reflect on the instruction they are providing. The process also provides increased opportunities for collaboration among general and special educators both in developing tasks and designing instructional and assessment accommodations. For all of these reasons, it appears that the use of performance assessments as part of large-scale programs for accountability offers important opportunities for enriching the education of students with disabilities.

Inclusion in the assessment program means increased attention to outcomes and increased scrutiny of the instruction provided to students with disabilities.

This is of critical importance if we are to view performance assessments as more than a fad approach to evaluating student performance. Rather, performance assessments appear to provide educators and policymakers the opportunity to define and assess educational outcomes that are more authentic and reflect higher-order thinking skills. At the same time, the assessments can link to classroom instruction and program improvement. Teacher and student involvement in the assessment process is increased, and there can be a clearer understanding of what is expected in the teaching/learning process.

Despite the positive opportunities offered by performance assessments, their use as part of large-scale accountability systems is troubled by issues related to scoring reliability and the resources required for administration, scoring, and professional development. Nonetheless, if the experiences of the five sites are any indications, there is a commitment and willingness to overcome the obstacles and maintain performance assessments as a featured part of an assessment program.

References

Baker, E., & Linn, R. (1993, Spring). The technical merit of performance assessments. *The CRESST Line*. Los Angeles: National Center for Research on Evaluation, Standards, & Student Testing, UCLA.

Brauen, M., O'Reilly, F., & Moore, M. (1994). *Issues and options in outcomes-based accountability for students with disabilities*. College Park: Center for Policy Options in Special Education, University of Maryland.

Council for Educational Development and Research. (1994). *Surveying the landscape of state educational assessment programs*. Washington, DC: Author.

Coutinho, M., & Malouf, D. (1993). Performance assessment and children with disabilities: Issues and possibilities. *TEACHING Exceptional Children*, 25(4), 62–67.

DeStefano, L. (1993). *The effects of standards and assessment on students in special education*. Minneapolis: National Center on Educational Outcomes, University of Minnesota.

Dietel, R. (1993). What works in performance assessment? *Evaluation Comment*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, UCLA.

Hasazi, S., Gordon, L., & Roe, C. (1985). Factors associated with the employment status of handicapped youth exiting high school from 1979-1983. *Exceptional Children, 51*, 455–469.

Hill, P., & Bonan, J. (1991). *Decentralization and accountability in public education*. Santa Monica, CA: RAND Corporation.

Jakweth, P., & Frey, W. (1992). *Including students with disabilities in educational evaluation: Implications of unique needs of learners*. Paper presented at the annual meeting of the American Evaluation Association, Seattle.

Kirst, M. (1990). *Accountability: Implications for state and local policymakers*. Washington, DC: U.S. Department of Education.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1993). *Interim report: The reliability of Vermont portfolio scores in the 1992-93 school year.* Santa Monica, CA: RAND Institute on Education and Training.

Koretz, D., Stecher, B., & Deibert, E. (1992). *The Vermont portfolio assessment program: Interim report on implementation and impact, 1991–92 school year.* Santa Monica, CA: RAND Corporation.

Linn, R., Baker, E., & Dunbar, S. (1991). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

McGrew, K., Thurlow, M., Shriner, J., & Spiegel, A. (1992). *Inclusion of students with disabilities in national and state data collection programs.* Minneapolis: National Center on Educational Outcomes, University of Minnesota.

McLaughlin, M., & Warren, S. (1994). Restructuring special education programs in local school districts: The tensions and the challenges. *Educational Leadership Review, 2*(1), 2–21.

National Center on Educational Outcomes, (NCEO). (1991). *Assessing educational outcomes: State activity and literature integration.* Minneapolis: University of Minnesota.

National Council on Disability. (1993). *Serving the nation's students with disabilities: Progress and prospects.* Washington, DC: Author.

Office of Technology Assessment. (1992). *Testing in America's schools. Asking the right questions.* Washington, DC: Author.

Olsen, K., & Massanari, C. (1991). *Special education program evaluation: What should states consider?* Lexington: Mid South Regional Resource Center, Interdisciplinary Human Development Institute, University of Kentucky.

Popham, W. (1993). Circumventing the high costs of authentic assessment. *Phi Delta Kappan, 74*(6), 470–473.

Poteet, J., Choate, J., & Stewart, S. (1993). Performance assessment and special education: Practices and prospects. *Focus on Exceptional Children, 26*(1).

Richards, C. (1988, March). Indicators and three types of monitoring systems: Implications for design. *Phi Delta Kappan*, 495–499.

Rothman, R. (1991, June 12). Coalition issues guide to help evaluate student-test systems. *Education Week*, p. 25.

Shavelson, R., Baxter, G., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21*(4), 21–27.

Smith, S. (1990). Individualized education programs in special education: From intent to acquiescence. *Exceptional Children, 57*, 6–14.

Thurlow, M. (1994). *National and state perspectives on performance assessment and students with disabilities.* Reston, VA: The ERIC Clearinghouse on Disabilities and Gifted Education, The Council for Exceptional Children.

Wagner, M. (1991). *The transition experience of youths with disabilities: A report from the National Longitudinal Transition Study.* Menlo Park, CA: SRI International.

Warren, S. (1993). *Assessment of and accountability for outcomes of students with disabilities: A survey of local school district practices.* Unpublished doctoral dissertation, University of Maryland at College Park.

Weber, L., & Zin, T. (1992, November). *The compatibility of individual and program evaluation: Is it reasonable to use data collected on individuals as a basis for determining the value of special education programs?* Paper presented at the annual meeting of the American Evaluation Association, Seattle.

The ERIC/OSEP Special Project

The ERIC/OSEP Special Project at The Council for Exceptional Children facilitates communication among researchers sponsored by the Office of Special Education Programs (OSEP) in the U.S. Department of Education, and it disseminates information about special education research to audiences involved in the development and delivery of special education services. These audiences include

- Teachers and related services professionals.
- Teacher trainers.
- Administrators.
- Policymakers.
- Researchers.

The activities of the ERIC/OSEP Special Project include tracking current research, planning and coordinating research conferences, and developing a variety of publications that synthesize or summarize recent research on critical issues and topics. Each year, the Special Project hosts a conference attended by research project directors sponsored by OSEP. Throughout the year, it holds research forums and work groups to bring together experts on emerging topics of interest. Focus groups representing the Special Project's audiences are held to inform both OSEP and the Special Project of audience information needs and to enhance the utility of publications produced by the Special Project. These publications include an annual directory of research projects as well as publications about current research efforts.

The ERIC/OSEP Special Project is funded under a three-party contract between The Council for Exceptional Children, the Office of Special Education Programs, and the Office of Educational Research and Improvement, U. S. Department of Education. Under this contract, OSEP funds the ERIC/OSEP Special Project, and OERI funds the ERIC Clearinghouse on Disabilities and Gifted Education. The ERIC Clearinghouse on Disabilities and Gifted Education is one of 16 clearinghouses of the Educational Resources Information Center (ERIC) system, which maintains a database of over 440,000 journal annotations and 340,000 document abstracts concerning education. The ERIC Clearinghouse on Disabilities and Gifted Education gathers and disseminates information on all disabilities and on giftedness across age levels.

A product of the ERIC/OSEP Special Project
The ERIC Clearinghouse on Disabilities and Gifted Education
The Council for Exceptional Children

53

ISBN 0-86586-250-8