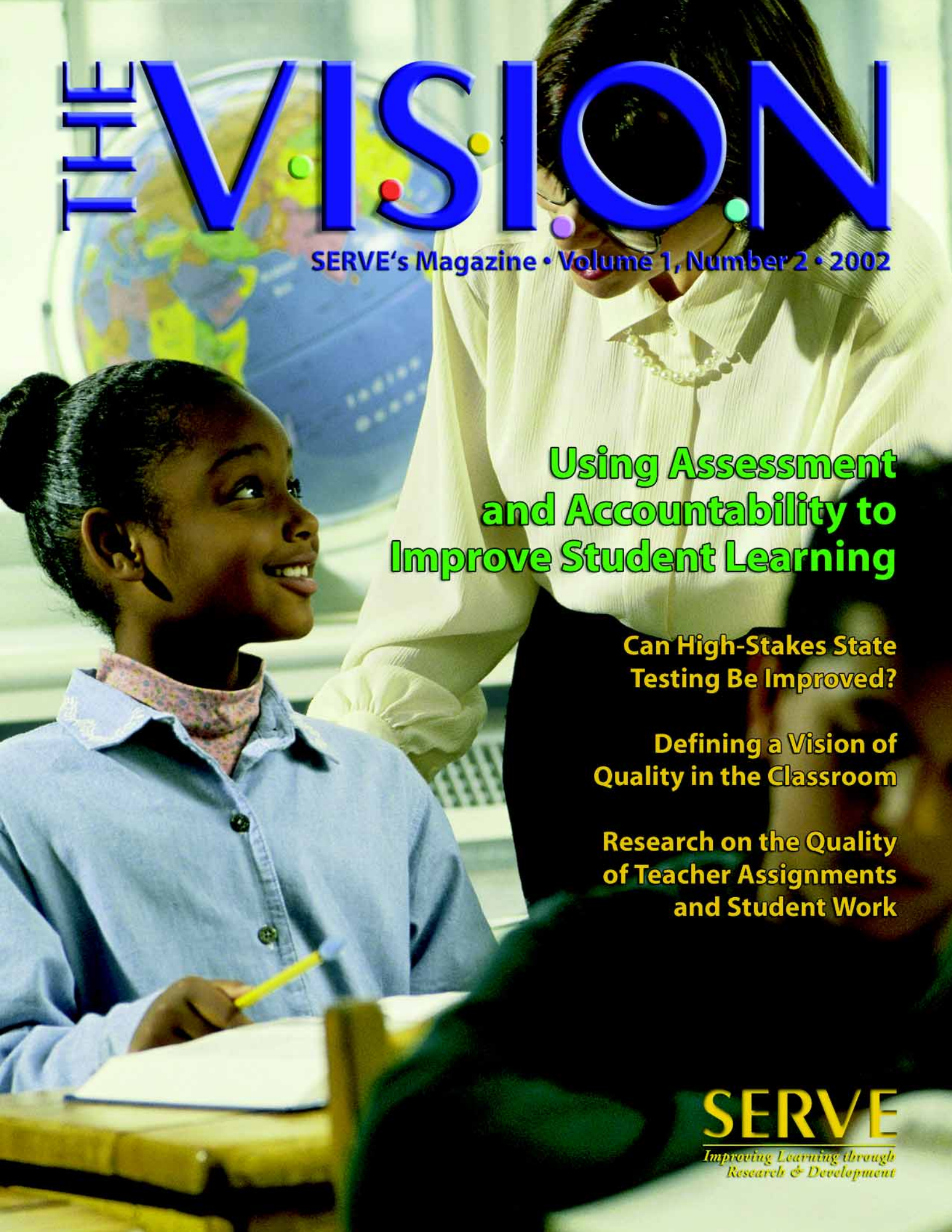


THE VISION

SERVE's Magazine • Volume 1, Number 2 • 2002

Using Assessment and Accountability to Improve Student Learning

Can High-Stakes State Testing Be Improved?

Defining a Vision of Quality in the Classroom

Research on the Quality of Teacher Assignments and Student Work

SERVE
Improving Learning through Research & Development

LETTER FROM THE EXECUTIVE DIRECTOR

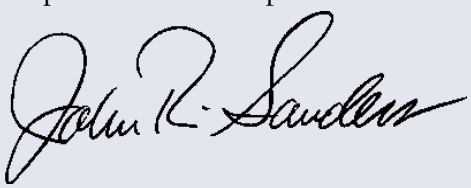
Dear Colleagues:

In this expanded issue of *The Vision*, we examine emerging thinking about assessment systems that leads to improved instructional practices, and we discuss the challenges educators and policymakers face when designing such systems. Accountability systems establish responsibility in an effort to raise educators' and students' performance to achieve a higher standard of quality. Designing systems that encourage improvement rather than just punish failure and that promote achieving higher standards rather than just higher test scores is a daunting and challenging task.

In our feature story, we report on the work of the Commission on Instructionally Supportive Assessments to define what a good state assessment system might look like and to offer recommendations for achieving systems that effectively improve teaching and learning. Included is an interview with the Commission's chairperson, Dr. James Popham, and a response to the Commission's recommendations by Mr. Lou Fabrizio, the Director of Accountability Services for the North Carolina Department of Public Instruction.

To effectively improve student achievement, district leaders need help in staying focused on improving the quality of learning opportunities their teachers provide students. In the article "Standards of Classroom Practice: Defining a Vision of Quality in the Classroom," we explore ways to support teachers in their efforts to provide high-quality, interesting, challenging, and purposeful learning experiences for students and discuss what some districts are currently doing in the Southeast to begin to achieve this vision of quality. Lindsay Clare Matsumara of CRESST continues the discussion of classroom quality by describing her organization's research aimed at developing a method for investigating the quality of students' learning environments by developing criteria for rating teachers' assignments.

In this issue's installment of our continuing series on relevant legislation, we look at the assessment requirements of the Leave No Child Behind Act and discuss the implications of those requirements for those designing state assessment systems. Finally, we return our focus to the Southeast with reports from our policy analysts in Florida and South Carolina and a description of our Southern States Seminar, where participants worked to create a regional response to school improvement efforts and state policy recommendations.



John R. Sanders, Ed.D.
Executive Director

THE VISION MAGAZINE

Volume 1, Number 2
2002

Produced by:
SERVE, The Regional Educational Laboratory for the Southeast

Associated with the School of Education, University of North Carolina at Greensboro

Edited by:
Wendy McColskey,
Director of Assessment, Accountability, and Standards, SERVE

Donna Nalley,
Publications Director, SERVE

Designed by:
Tracy Hamilton,
Publications Art Director, SERVE

Select photographs used in this magazine are from Comstock Klips, Corbis, Creatas, EyeWire Images, Image 100, and/or PhotoDisc.

SERVE North Carolina Office
800-755-3277 (Main Office)

SERVE Florida Office
800-352-6001

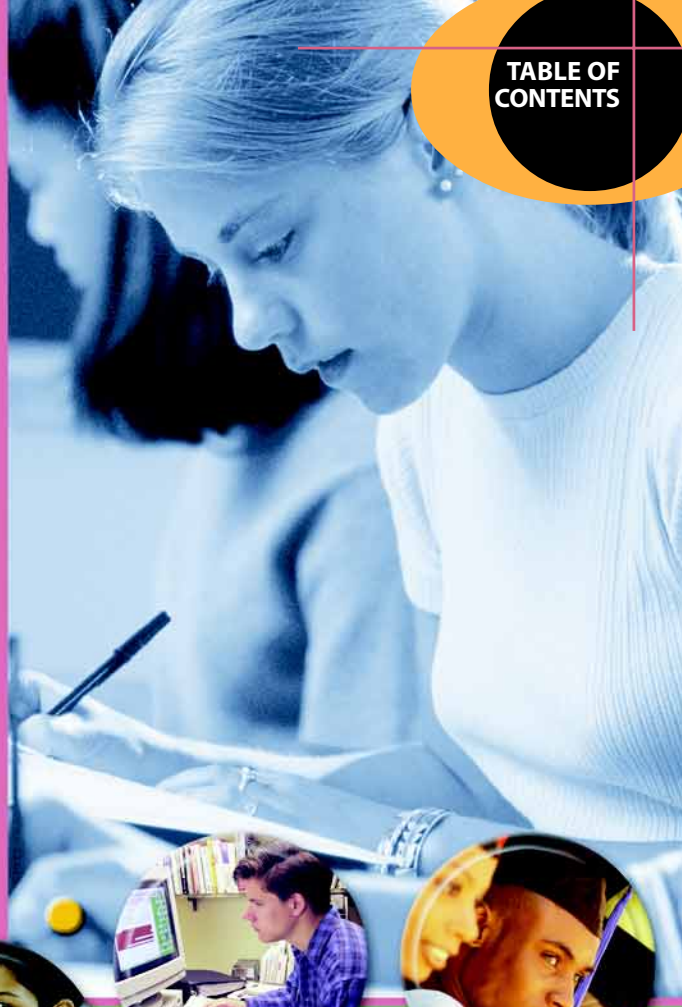
SERVE Georgia Office
800-659-3204

www.serve.org

The content of this publication does not necessarily reflect the views or policies of the Office of Educational Research and Improvement, U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This document was produced with funding from the Office of Educational Research and Improvement, U.S. Department of Education, under contract no. ED-01-CO-0015.

SERVE
Improving Learning through Research & Development

Contents



About SERVE

SERVE is an education organization with the mission to promote and support the continuous improvement of educational opportunities for all learners in the Southeast. To further this mission, SERVE engages in research and development that address education issues of critical importance to educators in the region and provides technical assistance to SEAs and LEAs that are striving for comprehensive school improvement. This critical research-to-practice linkage is supported by an experienced staff strategically located throughout the region. This staff is highly skilled in providing needs assessment services; conducting applied research in schools; and developing processes, products, and programs that inform educators and increase student achievement.

Can High-Stakes State Testing Be Improved?

By Dr. Wendy McColskey, SERVE Director of Assessment, Accountability, and Standards, and Nancy McMunn, SERVE Senior Program Specialist

Recently, a group of major educational professional associations took a proactive step in trying to influence the debate about the nature of state testing programs. These five associations (AASA, NAESP, NASSP, NEA, and NMSA—see sidebar for complete names and website contact information) were concerned that high-stakes state testing can have unintended negative consequences on the quality of teaching and learning. At the same time, they understood that some policymakers perceive educators as "running from accountability." Thus, this coalition of organizations turned to an independent commission of nationally recognized experts in assessment, curriculum, and instruction to attract the attention of state policymakers and urge them to examine their testing programs. Their goal was to engage the public in a debate about what a good state assessment system might look like.

The five associations are:

AASA—American Association of School Administrators (**www.aasa.org**)

NAESP—National Association of Elementary School Principals (**www.naesp.org**)

NASSP—National Association of Secondary School Principals (**www.nassp.org**)

NEA—National Education Association (**http://nea.org**)

NMSA—National Middle School Association (**http://nmsa.org**)

W. James Popham, a noted professor emeritus of the University of California and author of numerous articles and books on assessment, was invited to serve as the chair of the Commission on Instructionally Supportive Assessment, and he selected members of the Commission from a list of nominees submitted by the associations. The Commission functioned with complete autonomy from the convening organizations. The Commission's goal was to develop a set of recommendations that could potentially improve state-administered achievement tests so that they might serve the dual purposes of accountability (e.g., grading schools) and improvement of instruction, rather than just the single purpose of satisfying public demands for accountability.

The Commission's ten members released their report entitled, *Building Tests to Support Instruction and Accountability: A Guide for Policymakers*, in October 2001. The report, which is available on each of the associations' websites, advanced nine requirements for states to consider. The following statement accompanied the release of the report:

> *In the rush to implement testing systems, too few states have tests that are designed to help teachers improve teaching methods and curriculum or are useful in helping children learn higher-level thinking skills. As educators closest to America's students, we recognize that current tests fall short in the effort to improve teaching and learning. As state-mandated tests continue to be the measure of school quality, it is imperative that quality tests be implemented in a way that helps students achieve and acquire a love of learning.* (October 23, 2001 Press Release)

The nine requirements from the Commission's report are summarized at right. An interview with Dr. Popham follows to aid in conceptualizing the thinking behind the requirements. A response to the Commission's report from a state assessment director continues the dialogue. Finally, a set of discussion questions related to each of the nine requirements is offered.

Building Tests to Support Instruction and Accountability: A Guide for Policymakers

The report outlined the following nine requirements:

1. A state's content standards must be prioritized to support effective instruction and assessment.

This requirement recommends that states review and prioritize their content standards resulting in a high-priority set. The purpose of the prioritization is to identify a small number of content standards, suitable for large-scale assessment, that represent the most important skills and knowledge students need to learn in school.

2. A state's high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident.

This requirement builds on the assumption that better teaching and testing are more likely if educators clearly understand where they need to focus and test developers understand clearly what they are to test. Therefore, states should analyze their high-priority standards and identify what students must do and understand to demonstrate they have achieved standards. The analysis should result in relatively brief, educator-friendly descriptions of each high-priority standard's meaning. The high-priority standards and their descriptions need to be articulated across grade levels so that they build from grade-to-grade.

3. The results of a state's assessment of high-priority content standards should be reported standard-by-standard for each student, school, and district.

This requirement assumes that educators can do little to improve students' achievement without information about their performance on each high-priority content standard. In other words, standard-by-standard reporting of students' performance on state tests is critical to the success of standards-based educational reform. The Commission acknowledges that students will need to answer several items for each content standard assessed, which, in turn, means fewer content standards can be assessed by state tests given typical time constraints. They believe that having standard-by-standard classroom, school, or district-level information will enable educators to evaluate the effectiveness of instruction related to each standard and then improve instruction where needed. They point out that information based on just a few items is likely to be less reliable as a measure of an individual student's true knowledge and skills and suggest that teachers will have to bring additional sources of classroom-based information to their evaluation and intervention decisions for individual students.

4. A state must provide educators with optional classroom assessment procedures that can measure students' progress in attaining content standards not assessed by state tests.

This requirement recognizes that statewide tests measure a limited number of high-priority standards and that other important standards, not on the state tests, should receive instructional and assessment attention in the classroom. Therefore, states need to develop optional classroom assessments for these non-state-tested standards to support educators' efforts to teach a wide range of skills and knowledge. The Commission recommends that states conduct professional development for educators regarding how to best use optional assessments for instructional improvement and also how to design their own classroom assessments to measure students' progress in meeting state standards. States need to let

(continued)

educators know how results from classroom assessments can be reported alongside state test results to provide parents and policymakers with a complete picture of students' achievement on all the state's content standards.

5. A state must monitor the breadth of the curriculum to ensure that instructional attention is given to all content standards and subject areas, including those that are not assessed by state tests.

This requirement obliges states to support educators in curriculum coverage that extends beyond just those prioritized standards tested by the state. The Commission acknowledges that state assessments can inadvertently lead to a narrowing of the curriculum as educators work to ensure that students do well on state tests. A narrowed curriculum means dealing almost exclusively with content assessed on state tests. To prevent narrowing, this recommendation suggests monitoring curricular breadth at the state, district, and/or school levels using quantitative and/or qualitative methods.

6. A state must ensure that all students have the opportunity to demonstrate their achievement of state standards; consequently, it must provide well-designed assessments appropriate for a broad range of students, with accommodations and alternate methods of assessment available for students who need them.

This requirement obliges states to design state assessments or appropriate alternatives that will provide accurate and useful information for the teacher on the extent to which students with special needs have demonstrated the skills and knowledge described in the state content standards. This requirement is consistent with federal laws that obligate states to develop guidelines for districts about how all students participate in the assessment.

7. A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy the Standards for Educational and Psychological Testing and similar test-quality guidelines.

The Commission bases the need for this requirement on the belief that there exists widespread misunderstanding that high-quality achievement tests can be developed in two years or less. Recognizing that there is often pressure to produce tests quickly, they counter with the argument that these tests are far too important to be developed improperly. Experience shows that a minimum of three years is needed to develop a state test to assess high-priority standards in a way that promotes instructional improvement. The steps in developing a good test take time. They include prioritizing standards, determining skills/knowledge students must demonstrate for each standard to be tested, developing sufficient numbers of items, evaluating the items through small-scale pilots or other reviews, formal field-testing of all items, and assembling the final tests. Evidence regarding test quality, as called for in the *Standards for Educational and Psychological Testing,* is important to assemble after the test is developed.

8. A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments.

This requirement obliges states to educate policymakers and others about the need to provide professional development activities that will promote success in the use of the instructionally supportive assessment systems advocated by these requirements. Educators will need help in learning how to use information for instructional improvement and how to use or develop classroom assessments that supplement state tests.

9. A state should secure evidence that supports the ongoing improvement of its state assessments to ensure those assessments are (a) appropriate for the accountability purposes for which they are used, (b) appropriate for determining whether students have attained state standards, (c) appropriate for enhancing instruction, and (d) not the cause of negative consequences.

The tests that the Commission advances in this report represent a new generation of state tests that provide information for accountability as well as information for improving instruction. The new kind of assessment system is envisioned as a combination of state tests that focus on high-priority content standards combined with results from classroom assessments that focus on non-state-tested standards and together provide a more complete picture. The report suggests that states conduct independent evaluations to find out if state and classroom assessments are working together as envisioned. Other issues for independent review are the degree to which students had sufficient opportunity to learn the standards being tested, the degree to which tests are sensitive to differences in instructional quality, and the nature of any negative, unanticipated consequences of state tests, such as dramatically increased dropout rates.

FEATURE STORY INTERVIEW:

Interview with Dr. James Popham, Chair of the Commission on Instructionally Supportive Assessment

In an effort to understand the logic behind the nine requirements developed by the Commission on Instructionally Supportive Assessment for consideration by policymakers and educators, SERVE interviewed Dr. Popham in December 2001. The interview highlights are summarized below.

Purpose of the Report

Why did these groups feel the need to call this Commission together to develop these requirements?

Popham: The primary impetus was the likelihood that there would be a stringent requirement for testing across the U.S. based on pending (at that time) federal legislation calling for grade 3–8 testing. In addition, many educators feel the current approach to state testing has problems. Both of these factors contributed to the need for the report.

You mentioned in the report's foreword that all too often educators' legitimate concerns about state assessment are not heard; do you think this report will be heard?

Popham: If people dismiss the report because they think they already know what these organizations have to say, they will be making a mistake because what the professional organizations that sponsored the report promised the Commission was that they wouldn't interfere with what we came up with. They said that if we produced the report and did it without charge (so it would not look as if we were just doing this for fee), they would not interfere with the content. So the Commission did this report on the condition that whatever we came up with would not be changed by what they wanted, and the sponsoring organizations agreed and followed through. So it's really an independent commission that they convened.

Each of the organizations recommended a slate of people, but I selected the members. I did the choosing primarily on the grounds that these people would not only represent a diversity of gender, ethnicity, and so on, but also that those selected would be very, very knowledgeable regarding both assessment and instruction, and they were.

Understanding the Nine Requirements

Can you talk a little about the first of the nine requirements having to do with prioritizing the state's content standards and why this is the first requirement?

Popham: What people have to understand is that we have state curricular goals (i.e., content standards), but they've come up with an absolute plethora of content standards that realistically cannot be taught in the time teachers have available to them. In addition, because there are so many content standards, they absolutely cannot all be measured and certainly cannot be measured in such a way as to give feedback on a standard-by-standard basis. So the situation we have now with state testing is wrong. We've got to do something different to find out how schools are doing, something that gives educators a reasonable chance to improve the quality of instruction.

The hard reality of the current state-testing situation is that you're not measuring all the standards (because there are too many), teachers can't teach all the standards (because there are too many), and you can't give the citizens and teachers good feedback from test results about specifically what is going well and what isn't.

So one of the keys, according to your report, is that state content standards need to be more focused in terms of which ones a state is going to put its assessment dollars into (the first requirement)?

Popham: Exactly! What states can do is identify a group of curriculum experts (maybe the same people who came up with the content standards for that state) and ask them to put the content standards into three piles: the ones they think are absolutely essential, the ones they think are very desirable, and the ones they think are just desirable. Then, you have them look at just the absolutely essential ones and rank order them from the most important to least important. Then, this essential list is given to the assessment people, who given the amount of time they have for the test and the necessity to provide standard by standard test results, go as far down the list as they can with their test items, and, all of a sudden, you've got a very different kind of test.

What the Commission is saying is that states can't test all of the content standards that are sitting out there right now, and to pretend to do so is hypocrisy. Therefore, if you can't test them all, let us prioritize. If states prioritize and have a smaller number of content standards tested, then, of course, you have to ask about the content standards that don't make it on the state test. We have other requirements that directly emerge from the first requirement. However, our first and most important assumption was that a state test that does not help instruction ought not to be used, which led us to the first requirement, which, in turn, led to the others.

If you have a high-stakes test and do not identify with clarity what is to be assessed, except to say in a general sense that the test will assess state content standards, then teachers really don't know what will be coming at their students. That is the kind of uncertainty that breeds item-specific teaching (teaching to particular test items).

(continued)

The teacher may feel he or she can't win (i.e., I can't do a good job in teaching these students because I don't know what's going to be assessed). Out of fear of having students not do well, the teacher may design teaching guided by specific information about what kind of items will be on the state test. The current situation (of states not communicating exactly what content standards are most essential and assessed by items on state tests) is breeding a level of instructional ineffectiveness and instructional dishonesty that is very undesirable.

Implementation of Report Requirements

Do you think just publishing this report will result in changes in the design of state tests?

Popham: According to the folks at the NEA, there have been 14 states already that have expressed some interest in these requirements, and of the 14 that have expressed an interest, two or three might actually do something. So once you see the possibility of two or three states actually working on these requirements, it seems to me there might be some minds changing.

We developed a second report that contains sample language for a Request for Proposal (RFP) that would be issued by a state to create the kind of test that we're talking about. All a state has to do is get that RFP report and discover how to put language into the RFP that would force commercial test developers to develop a test in line with the recommendations of the Commission. The two documents together are crucial. We didn't want to come up with recommendations in the first report and then not give people the help on how to pull it off! If you look at some of the language in the RFP example, you will see that the kind of test the Commission envisions would be a very different test from the ones sitting out there now.

See *Illustrative Language for an RFP to Build Tests to Support Instruction and Accountability, 2001*, online from each convening association; e.g., *www.nea.org*.

Finally, I think we tried to describe specific rather than general requirements so that those in a state could easily assess where the state stands on these requirements. Either they do it or they don't. Let's say you have standard-by-standard reporting of your test results, where's the equivocation around that? Does the state do it or not? Does the state have a small set of prioritized content standards that it assesses on a state test, or does it develop a particular state test against a list of 200? Does the state supply optional classroom assessments for teachers; does it monitor the breadth of the curriculum in schools or districts so that students don't experience a narrowed curriculum that reflects just the focus of the state test; does it allow test developers a minimum of three years to produce statewide tests; does it ensure that educators receive professional development on how to use the results of instructionally supportive assessments? These are specific questions generated by the requirements that people can ask and answer of their state.

Most of the states in the Southeast already have state tests, but they probably don't produce the kind of standard-by-standard reporting your Commission is suggesting. Is that accurate?

Popham: That is one of the problems we will face with the new ESEA legislation. One of the easiest things to do is to say we've got tests in place, but, unfortunately, the tests may not be as good as they could be. The states may have state-developed tests aligned with their content standards, but they still don't meet the requirements the Commission has outlined. They don't have standard-by-standard reporting, which we believe is pivotal if you are going to have standards-based assessment that improves the quality of instruction. If you cannot find out from the test which standards kids are getting and which ones they're not, you can't help them improve. The Southeast has historically moved more rapidly in state testing than other places, but the kinds of tests they are using right now are not the kinds of tests that the Commission envisions.

Would you envision it costing a state more money to implement some of the requirements that go beyond the actual test development process (e.g., optional classroom assessments and ongoing evidence of effectiveness of state tests)?

Popham: Those kinds of things should be done in any state department and thus, ideally, should not represent new costs. The idea of continually evaluating to see whether your testing program is working (Requirement Nine) is just sensible management. As far as producing classroom assessment examples for teachers, a number of states are doing that right now based on the realization that they have to produce some of the tools necessary for teachers to use in the classroom. The big expense for some states will be building new tests, and, ideally, there will be some federal dollars there.

What can you say about the state and district relationship as regards responsibilities for student assessment?

Popham: If we have a statewide test that measures a smaller number of content standards as we're suggesting, and does so honestly, telling teachers the ones that will be measured, then there's going to be a tendency for curricular reduction or narrowing. So what we're suggesting to counteract this response at the local level is to 1) come up with some way of monitoring the extent to which the breadth of the curriculum is being taught (Requirement Five) and 2) to supply some tools that would make it easier for teachers to pursue those content standards not assessed at the state level (Requirement Four). These functions could be designated as state or district responsibilities.

District-developed assessments would be optional. We have an obligation especially with the new federal law to say

clearly to districts: here are the content standards that we (the state) really want you to teach, and they will be assessed by our state tests, and we will give you standard by standard reporting. However, if you want to go beyond our state assessment focus and assess some other important standards on your own, go for it. At least that way, the district will be able to plan and use their assessment development resources wisely to complement and build on what the state is doing with its assessment program. If I were running a district, I would want to see how we were doing on the small set of essential standards tested by the state, and I'd want to see how we were doing on the other important standards not assessed by the state, but that should be a local call.

Is there an organization that intends to monitor how states begin to consider these requirements?

Popham: The Commission's report was released at a press conference at the National Press Club on October 23, 2001, in Washington D.C. At that time, NEA announced that one year after the enactment of the federal law (ESEA), it would begin to annually monitor the extent to which states were creating tests consistent with these requirements. So there is going to be a systematic annual appraisal of states by the professional associations that convened the Commission's report.

Members of the Commission

Commission Chair: W. James Popham, Professor Emeritus at the UCLA Graduate School of Education and Information Studies and a past president of the American Educational Research Association

Eva L. Baker, Professor of Educational Psychology and Social Research Methods at the UCLA Graduate School of Education and Information Studies and the Co-Director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

David Berliner, Regents' Professor, School of Education at Arizona State University

Carol Camp Yeakly, Professor of Urban Politics and Policy at the Curry School of Education, University of Virginia

James W. Pellegrino, Liberal Arts and Sciences Distinguished Professor of Cognitive Psychology and Distinguished Professor of Education, University of Illinois at Chicago

Rachel Quenemoen, Senior Fellow for Technical Assistance and Research at the National Center on Education Outcomes, University of Minnesota

Flora V. Rodriguez-Brown, Professor of Curriculum and Instruction/Reading, Writing, and Literacy, University of Illinois at Chicago

Paul Sandifer, former Director of the South Carolina Department of Education's Office of Student Performance Assessment

Stephen G. Sireci, Associate Professor in the Center for Education Assessment, School of Education, University of Massachusetts

Martha L. Thurlow, Director for the National Center on Education Outcomes at the University of Minnesota

FEATURE STORY RESPONSE:

A State Assessment Director Responds to the Suggested Requirements

By Louis M. Fabrizio
(These comments do not necessarily reflect the opinions of the North Carolina Department of Public Instruction or SERVE.)

Editor's Note: *North Carolina is often referred to as a model for what state testing programs should be doing. The state tests were designed primarily for school accountability and, more recently, individual student accountability purposes. Because of his experiences in North Carolina, SERVE asked Lou Fabrizio, the Director of Accountability Services for the North Carolina Department of Public Instruction, for his reactions to the Commission's recommendations for states.*

The report, *Building Tests to Support Instruction and Accountability: A Guide for Policymakers,* is concise and easy to read, but somewhat idealistic. The authors, all well regarded and nationally known, have pieced together nine requirements for a new generation of state achievement tests, which are hard to argue against. However, the requirements seem to ignore the fact that no single test can do everything, as the comparison below of characteristics of tests designed for different purposes illustrates. It would seem that the authors of this new report believe that a single test can serve both purposes.

Two areas that resonate positively, based on my experiences in North Carolina, specify that test developers should have a minimum of three years to produce statewide tests (**Requirement 7**) and that states should ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments (**Requirement 8**). Obtaining the resources and staff to provide the professional development has been very difficult for states in the past, however, and professional development is an easy target when budget cuts are needed.

Requiring states to prioritize content standards (**Requirement 1**) is a good idea in theory, but even the Commission's authors caution that schools and teachers may resort to focusing their instruction on the high-priority standards and not address all standards. Their response to schools that might not teach all standards is **Requirements 4 and 5**. These requirements suggest that states provide educators with optional classroom assessments for the lesser priority standards and monitor the breadth of the implemented curriculum to ensure that instructional attention is given to all content standards and subject areas. Again, these requirements sound

Comparison of Assessments Designed for Different Goals

Assessment Designed for Measurement
Valid Reliable
Objective Cost-efficient Time-efficient
Centrally mandated
Widely applicable Centrally processed
Multiple-choice
Machine-scorable Delayed feedback
Used independently
Formal
Producing stable scores
Results designed for external user

good in theory, but I question the practicality of offering "optional" classroom assessments (**Requirement 4**) or asking the state to monitor the breadth of the curriculum (**Requirement 5**). Specifically, in **Requirement 5**, the authors recommend "monitoring of curricular breadth at these same three levels [state, district, and school] using quantitative and/or qualitative methods that states and school districts develop." I believe this recommendation will be difficult to implement. Many state departments of education are facing budget and staff shortages. Securing funds to hire staff to do the monitoring could be an issue.

Requirement 3 suggests that states report test results "standard-by-standard" for each student, school, and state, but in their discussion, the authors caution about the unreliability of making these decisions based on just a few items. The score reporting is further compromised when you consider that the tests would only measure high-priority standards and not represent sampling of all standards.

The solution of combining ongoing classroom assessment results with the results of state assessments is problematic due to measurement difficulties in combining results from non-mandated classroom assessments throughout the school year with the results of the statewide assessments across the various schools. Moreover, if the classroom assessments are optional, what happens if a school system decides not to use them?

Requirement 6 is a necessary requirement in light of federal statutes, but one that continues to be of major concern to states as they implement alternate assessments and offer accommodations for students with disabilities.

Although many essential topics were covered in the report, some important topics were not mentioned at all or were discussed only briefly. For example, field testing was mentioned very briefly in **Requirement 7**, but continually developing tests to meet these requirements would be both time consuming and expensive for schools as well as state departments of education. Furthermore, the significant issue of releasing test questions was not addressed. There was not much discussion of the timing of the tests. If the tests are administered at the end of the year, which I assume from my reading of the material, the test format becomes an issue. Are the tests solely multiple-choice, or do they contain constructed response items as well? If the tests contain constructed response items, the amount of testing time increases, and the number of items able to be administered in a typical testing time period decreases. The tests would need to be administered a month or two prior to the end of the school year to ensure that the items can be scored and the results reported to the schools before summer vacation.

At the time the authors wrote this report, there was no way to know what the results of the new reauthorization of the Elementary and Secondary Education Act (ESEA) would involve. That legislation now mandates annual testing in reading and mathematics in grades 3–8 beginning in the 2005–2006 school year and in the area of science beginning in 2007–2008. It will be interesting to see what impact this report has on states' efforts to meet the requirements of the new ESEA. There seems to be a trade-off between what is tested, how many items per standard will be included, and how long, in terms of minutes and/or days, students will need to take these various tests. The most ambitious part of the new ESEA, however, is the stipulation that all students must be proficient on state assessments by the year 2014. Finding a guide that will help us get to that goal is what we really need.

Assessment Designed for Instruction
Quality judged by effect on instruction
Design determined by instructional goals
Instructional raison d'etre Teacher-mandated
Adapted to local context
Test task of instructional value
Locally scorable Immediate feedback
Used with other information
Informal
Results subject to short-term change
Meaningful to students

Source: *Cole, N. S. (1987).* A realist's appraisal of the prospects for unifying instruction and assessment. *New York: ETS Invitational Conference.*

FEATURE STORY DISCUSSION:

SERVE Offers Discussion Questions Around the Nine Requirements

A state's content standards must be prioritized to support effective instruction and assessment.

- What process will be used to prioritize standards?
- Are current state standards good enough to use as a starting point for prioritization, or will standards need to be re-written?
- Should the enhanced descriptions of each high-priority standard (in Requirement 2) be developed before the prioritization process so that it is clearer what the standards mean before they are prioritized?

A state's high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident.

- How should prioritized standards be explained to educators so that they understand these are prioritized for the purpose of focusing large-scale testing efforts?
- If high-priority standards are identified for which it is difficult to develop easily scored paper-and-pencil items (perhaps an extensive performance task is more appropriate), should states still develop items for a state test even though the item types used might lead to inappropriate assessment practices in classrooms (as teachers mimic state test items in their classroom assessment)?
- Where do performance standards fit into the picture?

The results of a state's assessment of high-priority content standards should be reported standard-by-standard for each student, school, and district.

- If individual student results are not very reliable standard-by-standard, should they even be reported?
- What kind of information or training would a state have to provide to ensure that individual student standard-by-standard results were not misused?

A state must provide educators with *optional* classroom assessment procedures that can measure students' progress in attaining content standards not assessed by state tests.

- Who will develop the classroom assessments to measure standards that are not assessed by the state? Who will pilot and validate the assessments as being good for classroom use?
- Will classroom assessments be developed only for grade levels and content areas tested at the state level?
- How will educator expertise in classroom assessment be continuously developed in the state?

A state must monitor the breadth of the curriculum to ensure that instructional attention is given to all content standards and subject areas, including those that are not assessed by state tests.

- Who will do this monitoring, and what kinds of resources would be needed?
- What will happen if a school or district is found to be narrowing the curriculum?

A state must ensure that all students have the opportunity to demonstrate their achievement of state standards; consequently, it must provide well-designed assessments appropriate for a broad range of students, with accommodations and alternate methods of assessment available for students who need them.

- What is the process for developing standard-by-standard reporting that could be reliably reported out for individuals with special needs?

A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy the Standards for Educational and Psychological Testing and similar test-quality guidelines.

- If new forms of tests are developed each year from items found in a large item bank, should each year's tests be formally field-tested?
- Prior to assembling the final test, is it important to review test drafts in a subject area across the grade progression tested (grades 3-8) so that questions about relative grade appropriateness of items and adequacy of representation of most important concepts to master in each grade level can be addressed in the context of the whole set of tests?
- Should evidence as to the validity of the tests be provided for each annual form of the test?

A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments.

- Whose job responsibility in states or districts will it be to ensure all educators (principals, district staff, teachers) are assessment-literate?
- How should an awareness of the need for improvement in assessment literacy be developed in a state or district? What is the state role? What is the district role? University role?
- Given all the competing demands for the small amount of professional development time built into the school schedule (e.g., to support district and school improvement initiatives that change each year), how can states invite principals and teachers to become more assessment-literate without also addressing the issue of time?

A state should secure evidence that supports the ongoing improvement of its state assessments to ensure those assessments are (a) appropriate for the accountability purposes for which they are used, (b) appropriate for determining whether students have attained state standards, (c) appropriate for enhancing instruction, and (d) not the cause of negative consequences.

- Who is the audience for the evidence on how the assessment system can be improved (e.g., an external oversight committee), and how will the results be discussed in public forums with educators who are on the receiving end of state assessment systems so that a well-functioning feedback loop from the classroom to the policymakers and state department is ensured?
- Should information regarding how state tests are being used for individual student promotion/retention decisions be collected from districts with feedback given to districts about the appropriateness or inappropriateness of their use of state test results for this purpose (given the warnings from measurement experts that a single test should not be used for promotion decisions)?

FEATURE STORY RELATED WORK:

CRESST Defines Standards for State Accountability Systems

By Dr. Joan Herman, Co-Director, CRESST, UCLA

Editor's Note: *The work of the Commission reported on earlier is one attempt to define criteria or guidelines against which states can evaluate their testing programs. The work of the Commission focused on state tests, how they are developed, and how student assessment information from various levels might be brought together to improve instruction. Others in the field are working on standards for state accountability systems. State accountability systems are typically designed to create incentives for students, schools, and districts to focus on improving student achievement. Most states hope to bring some or all students to proficient levels of performance. The measure of proficiency that students or schools are held to is defined differently across states. Definitions of Adequate Yearly Progress are used to identify schools and districts in need of improvement. In other words, state accountability systems typically involve defining and measuring student outcomes, making judgments about the adequacy of the student achievement results, and attaching consequences to the results achieved by students, schools, and districts. The work on standards for accountability systems described below starts from the point of view that accountability is currently a major purpose of state testing.*

CRESST, the National Center for Research on Evaluation, Standards, and Student Testing, in partnership with the Consortium for Policy Research in Education (CPRE), with the Education Commission of the States (ECS), and with advice and review from numerous colleagues in research and practice, offers Standards for Educational Accountability Systems. These standards are intended to provide guidance to states and districts in conducting self-reviews of their own systems and to delineate criteria by which developing accountability systems can be judged. The Standards for Educational Accountability Systems represent compiled knowledge developed from sources including the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), research findings on testing and accountability systems, and studies of best practices.

Because experience with accountability systems is still developing, the standards are intended to help evaluate existing systems and to guide the design of improved procedures. The standards strongly endorse each state's responsibility to conduct continuing evaluation of its own accountability system. It is not possible at this stage in the development of accountability systems to know in advance how every element of an accountability system will actually operate in practice or what effects it will produce. Evaluations—conducted in-house or by universities, external organizations, or teams of experts—are essential if states are going to learn systematically from one another and the nation is going to judge the effectiveness of its efforts for children. Evaluation results will be essential to the continuing improvement of testing programs and accountability provisions.

In sum, the standards represent models of practice derived from three perspectives: research knowledge, practical experience, and ethical considerations. They should be conceived of as targets for state and local systems and as criteria to judge proposed models of accountability development. (It should be understood that tests included in an accountability system should meet the Standards for Educational and Psychological Testing.) What is highlighted here are criteria that apply especially to accountability systems. It is likely that additional standards will subsequently be developed based on reported evaluations of accountability system effects.

The Standards were developed by Dr. Eva Baker, Dr. Robert Linn, Dr. Joan Herman, and Dr. Daniel Koretz, and are organized into five categories:

1. Standards on System Components

- Accountability expectations should be made public and understandable for all participants in the system.
- Accountability systems should employ different types of data from multiple sources.
- Accountability systems should include data elements that allow for interpretations of student, institution, and administrative performance.
- Accountability systems should include the performance of all students, including subgroups that historically have been difficult to assess.
- The weighting of elements in the system, different test content, and different information sources should be made explicit.
- Rules for determining adequate progress of schools and individuals should be developed to avoid erroneous judgments attributable to fluctuations of the student population or errors in measurement.

2. Testing Standards

- Decisions about individual students should not be made on the basis of a single test.
- Multiple test forms should be used when there are repeated administrations of an assessment.
- The validity of measures that have been administered as part of an accountability system should be documented for the various purposes of the system.
- If tests are to help improve system performance, there should be information provided to document that test results are modifiable by quality instruction and student effort.

- If test data are used as a basis of rewards or sanctions, evidence of technical quality of the measures and error rates associated with misclassification of individuals or institutions should be published.
- Evidence of test validity for students with different language backgrounds should be made publicly available.
- Evidence of test validity for children with disabilities should be made publicly available.
- If tests are claimed to measure content and performance standards, analyses should document the relationship between the items and specific standards or sets of standards.

3. Stakes

- Stakes for accountability systems should apply to adults and students.
- Incentives and sanctions should be coordinated for adults and students to support system goals.
- Appeal procedures should be available to contest rewards and sanctions.
- Stakes for results and their phase-in schedule should be made explicit at the outset of the implementation of the system.
- Accountability systems should begin with broad, diffuse stakes and move to specific consequences for individuals and institutions as the system aligns.

4. Public Reporting Formats

- System results should be made broadly available to the press, with sufficient time for reasonable analysis and with clear explanations of legitimate and potential illegitimate interpretations of results.
- Reports to districts and schools should promote appropriate interpretations and use of results by including multiple indicators of performance, error estimates, and performance by subgroup.

5. Evaluation

- Longitudinal studies should be planned, implemented, and reported, evaluating effects of the accountability program. Minimally, questions should determine the degree to which the system
  - Builds capacity of staff
  - Affects resource allocation
  - Supports high-quality instruction
  - Promotes student equity access to education
  - Minimizes corruption
  - Affects teacher quality, recruitment, and retention
  - Produces unanticipated outcomes
- The validity of test-based inferences should be subject to ongoing evaluation. In particular, evaluation should address
  - Aggregate gains in performance over time
  - Impact on identifiable student and personnel groups

NOTE: A more comprehensive version of the standards, including comments, will be published as a forthcoming CRESST policy brief. See www.cse.ucla.edu.

STANDARDS OF CLASSROOM PRACTICE:

DEFINING A Vision OF Quality IN THE Classroom

By Nancy McMunn, SERVE Senior Program Specialist and
Dr. Wendy McColskey, SERVE Director of Assessment, Accountability, and Standards

You can hear the frustration in their voices when teachers say: *You want us to increase test scores, but you don't help us figure out how to do it.* Or when a high school math teacher says: *It's so frustrating. I teach this Algebra I concept over and over, and they just don't get it. How will they pass the state end-of-course test? What can I do?* Or when a veteran teacher says: *I know I'm not teaching in ways that will help my students with their long-term development as learners, but I feel pressured to "drill and kill" for them to do well on state tests because that's what I'm going to be judged on.*

District and school leaders are struggling to find ways to funnel teachers' frustrations with pressure for higher test scores into a focus on improving the quality of learning opportunities they provide their students. In continuously improving the quality of learning experiences they provide students, teachers need ideas, tools, and feedback from outside experts. They also need help in using each other as resources and critics in designing more interesting, challenging, and purposeful learning experiences for students and in examining the quality of learning that results. Outside or external help, in the form of books, Web resources, purchased materials, training, and consultants, is often available. Internal support, in the form of a highly professional working environment that provides time for the continuous improvement of teaching to enhance student motivation and learning, is often lacking.

The Challenges of Accountability

In the early 1990s, concerns that American students needed to improve their thinking and problem-solving skills for the United States to remain internationally competitive were reflected in the student outcomes described in national and state standards documents. These standards documents (however overloaded or ill-defined in some cases) were commendable attempts to articulate a new vision for student performance. The standards implied a need for more cognitively engaging and rigorous instruction and assessment in classrooms. In the late 1990s, as strong state accountability policies were implemented in some states, the goal of getting more students to achieve grade-level proficiency (sometimes inappropriately defined by a single score on a state test) may have inadvertently supplanted (rather than supplemented) the goal of providing more rigorous and challenging instruction to all students.

The crux of the problem we now face with external accountability and high-stakes testing, according to Tony Wagner, co-director of the Change Leadership Group at Harvard University's Graduate School of Education, is that "paper-and-pencil, computer-scored state tests do not begin to assess the competencies that large majorities of adults agree are essential today: the ability to comprehend difficult reading material and apply information to the solution of complex problems; the ability to write and speak clearly and thoughtfully; the ability to understand mathematical data and use math and technology as problem-solving tools; the ability to work effectively in teams; and finally, respect for others and an understanding of our roles as citizens."

In other words, the competencies assessed by state tests can't represent the total vision schools have for students' development. One implication of Wagner's statement is that if students are going to develop competencies that go beyond what can be assessed on a state test, teachers must provide high-quality learning opportunities reflecting the larger vision. The kinds of work one would see students doing in schools and districts focused on developing the skills that Wagner described are very different from the kinds of work one would observe students doing in schools and districts that focus exclusively on preparing students to do well on state tests.

For example, students who only understand reading comprehension to mean getting the right answers to multiple-choice questions about a passage or book (as found on state tests) will be much less able to face the challenges of difficult high school and college courses than students who have come to understand reading comprehension in a broader sense (e.g., summarizing, explaining, discussing, interpreting, evaluating, and, in general, making sense of what they read). Students come to these different understandings

(continued)

External accountability demands for better test scores arrive at the doors of district leaders. Some districts leverage external accountability demands into strategies that build school capacity for a higher level of professionalism and commitment to the quality of education provided students. They realize that passing along more edicts to schools will not produce higher levels of learning.

Internal accountability means "how a school system hires, evaluates, and supports its staff; how it makes decisions; how it acquires and uses the best knowledge; how it evaluates its own functioning; and how it safeguards student welfare" (Darling-Hammond, 1997, p. 245). Successful districts use external accountability as an opportunity to build internal accountability.

Reducing the Achievement Gap through Challenging Instructional Opportunities

Consider the following research findings that Linda Darling-Hammond reports in her 1997 book, *The Right to Learn:*

- When students of similar achievement levels are exposed to more and less challenging material, those given the richer curriculum systematically outperform those placed in less challenging classes (Alexander & McDill, 1976; Oakes, 1985; Gamoran & Berends, 1987).
- Teacher interaction with students in lower-track classes is less motivating and less supportive and also less demanding of higher-order reasoning and responses (Good & Brophy, 1986). The interactions are also less academically oriented and more likely to focus on criticisms of students' behavior, especially for minority students (Eckstrom & Villegas, 1991; Oakes, 1985). Learning tasks are less engaging and students are less engaged (Rosenbaum, 1976).

A substantial body of research has found that much of the difference in school achievement among students is due to the effects of substantially different school opportunities and, in particular, greatly disparate access to high-quality teachers and teaching. Although most research on the relationship between learning opportunities and outcomes is correlational, experimental studies offer strong evidence that what students learn is substantially a function of the opportunities they are provided.… It is becoming clear that differences in teacher expertise are a major reason for the difference in learning opportunities across schools and classrooms (Darling-Hammond, 1997, p. 271).

of reading comprehension, in part, from the kinds of work they are asked to do in school. Although state test scores provide a glimpse into what students can do, they reveal little about the quality of the learning opportunities students routinely experience over time in a particular district or school.

Low organizational expectations for teachers' classroom teaching make providing rigorous learning opportunities to all students even more challenging. As Elmore puts it, the perennial critique of American education is that "teaching and learning is in its most common form emotionally flat and intellectually undemanding and unengaging. Every school can point to its energetic, engaged, and effective teachers.… We regularly honor and deify these pedagogical geniuses. But these exceptions are not the rule. For the most part, we regard inspired and demanding teaching as an individual trait of teachers, much like hair color or shoe size, rather than as a professional norm, or an expectation that might apply to any teacher" (Elmore, 1996, p. 299).

To the extent that continuous improvement toward high-quality teaching is not the norm in many schools, then district leaders face a challenge in improving the quality of classroom learning opportunities provided students. Leadership and support to help teachers take responsibility for the effectiveness and quality of learning experiences they design and provide is critical.

High-Quality Learning Opportunities for Students

"If academic outcomes are to change, aggressive action must be taken to change the caliber of learning opportunities students encounter" (Darling-Hammond, 1997, p. 277).

For districts, improving student outcomes must involve continuous work on the quality of learning opportunities provided. Without explicit discussions about what constitutes high-quality classrooms, it is easy for schools to lose focus and get caught up in implementing a myriad of strategies that, although important, may have little impact on the quality of classroom instruction (such as changing schedules, buying technology, involving more parents and businesses, offering more outside tutoring, etc.).

Dr. Phillip Schlechty, founder and CEO of the Center for Leadership in School Reform in Louisville, Kentucky, argues convincingly that schools must focus on what they do to produce results rather than getting preoccupied with obtaining certain results. In other words, "schools do not produce test scores; students produce test scores." Schools should focus on the quality of the "knowledge work" in which they ask students to engage. He has coined the term "working on the work" (WOW) to describe the importance of teachers creating and constantly improving upon the schoolwork they assign to their students.

For teachers to design high-quality work and experiences for their students, they need to be able to think and talk about the work they give students in terms of how it meets certain criteria. Schlechty defines ten qualities of work (which he calls the WOW Framework) that teachers might use in their professional dialogues and reflections about the instructional program:

1. Content and substance
2. Organization of knowledge
3. Product focus
4. Clear and compelling product standards
5. Protection from adverse consequences for initial failures
6. Affirmation of the significance of performance
7. Affiliation
8. Novelty and variety
9. Choice
10. Authenticity

Schlechty is not suggesting that these attributes need to be applied to every assignment given to students, but that teachers need to be supported as they begin to apply these criteria. His organization has worked with many districts in trying to engage schools in applying these attributes to think about the work students do.

The Case of Community District #2

Richard Elmore, a highly regarded education researcher and writer, has studied the progress of Community District #2 in New York, a district that has engaged in a long-term process of system-wide instructional improvement. He contrasts this district's approach to other districts that tweak curriculum and instruction through the introduction of new programs and projects that have short life spans. According to Elmore, "this scattershot approach—popular with schools trying to keep up with the latest fads—may be 'the professional equivalent of yo-yo dieting'" (p. 157). Embedded in the district's continuous improvement philosophy is the assumption that the district must have a well-thought-out strategy for how it will influence the quality of classroom instruction. Elmore suggests that districts that are most likely to succeed in responding to the external pressure for student performance (state accountability) are those that use this pressure as an opportunity to build internal improvement processes aimed at the classroom.

> Professional development for teachers should be school-based, preferably embedded in instructional efforts through collaborative analysis of student work. This is contrary to most traditional professional development, such as courses leading to certificates or degrees but unrelated to the specific needs of the school, quick-fix workshops that do not offer consistent feedback, or professional development offered by external trainers to help teachers adopt specific programs. The National Commission on Teaching and America's Future recommends teachers "develop professional discourse around problems of practice" as a central component of professional development. What is needed, the Commission says, is replacing the isolation of teaching with "forums in which teaching and learning can be discussed and analyzed, and where serious examination of practice, its outcomes, and its alternatives is possible" (Lewis, 2001, p. 22).
>
> — *From a report on Using Research to Improve Education for Low Income and Minority Students*

Community District #2 began its focus on improving the quality of instruction in the late 1980s with teacher study groups, seminars, professional development on the teaching of literacy, and identification of models of good classroom practice (i.e., benchmarking). As a result of this work, the district began to develop "implicit standards of practice" that were commonly held expectations about what good teaching and learning in literacy was supposed to look like. They created a common language and mental model for good teaching in literacy. As time progressed, the district shifted into a more explicit focus on standards of practice in classrooms and on standards for student performance. In this district, explicit beliefs about classroom practice "grew out of the work" on improving instruction.

Quality Learning Opportunities: Continuous Improvement Takes Time

Studies show that teachers think they are providing challenging learning opportunities to students to a greater extent than they actually are. One study of biology teachers' assessment practices showed that although teachers reported broad thinking and problem-solving goals for their students, their assessment practices, in reality, focused on recognition and recall kinds of skills (Bol & Strage, 1996). Even in college classrooms, those who have studied assessment practices have found that teachers tend to think they are teaching to higher-order thinking goals, but when student assignments are examined, higher-order thinking goals tend to be weakly represented (Angelo & Cross, 1993). Similarly, Spillane and Zeuli (1999), in a study of 24 mathematics teachers who said they were implementing mathematics reform as outlined by the NCTM standards, found that only four of those teachers were giving students work that reflected the broader problem-solving and reasoning goals one would expect to see.

Equalizing the quality of learning opportunities does not come about by just putting out some literature for teachers or conducting a workshop. Surprisingly, it may not come about by adopting an innovative curriculum either because good instructional materials can be used poorly. In a presentation at a SERVE conference, Dr. Sam Stringfield, from the Center for Research on the Education of Students Placed At Risk, described the turnaround of a low-performing, high-poverty urban school (for more information, go to *www.csos.jhu.edu/crespar/reports/report20.pdf*). He reported that the school adopted the rigorous curriculum of a well-regarded, elite private school but phased it in slowly (one grade level at a time), with a staff person assigned to work with teachers on implementation and also providing feedback. In other words, experience suggests that the hard work of improving the quality of learning experiences provided students happens teacher by teacher in a professional work environment where hard questions about practice are discussed and individualized feedback to teachers is common practice.

Defining Standards of Classroom Practice

At some point in the process of trying to improve learning opportunities across schools and teachers in a district, educators must define, research, discuss, and refine the concept of "quality learning opportunities." For example, many districts, as they have taken on the goal of improved literacy instruction over a sustained period of time, have realized a need to define, as best they could, what constitutes good instructional practice in literacy. It's the "vision question": what kind of classroom learning environments do we envision for this school or district that will engender the highest levels of learning for all our students?

When experienced educators come together, a common vision for good classroom practice emerges. SERVE has been working with a consortium of districts for the past five years called SERVE-Leads. Twice a year, SERVE convenes a meeting for district leadership teams to discuss, share, and work on issues related to improving student learning. At a meeting in February 2001, SERVE staff members spent a day asking each district team to think about what is important in establishing quality classroom learning environments. Although each district's list was worded and organized differently, there were strong commonalities. For example, all the teams indicated that engaging students in interesting and challenging dialogue and cognitive work was central to achieving quality in the classroom.

The following Standards of Classroom Practice represent a synthesis of what SERVE-Leads educators believed were the characteristics of a powerful learning environment. The list is short, by design, in that participants were asked to come to agreement on five-to-seven compelling characteristics of a high-quality learning experience. The statements are intended to encourage professional dialogue and self-assessment.

(continued)

1. Instruction is organized and delivered in well-planned units that focus on clear learning goals and expectations for student learning.
2. Work assigned to students is purposeful, frequently cognitively challenging, and results in useful feedback from the teacher to help students improve.
3. Daily instructional interactions between teacher and student are structured to promote the development of key thinking and communication skills.
4. Persistence and effort on the part of students are nurtured.
5. Strategies are implemented to help students develop responsibility for their progress toward stated goals through self-assessment of the quality of their own work and thinking.
6. Meaningful social and interpersonal relationships are established that honor individual differences and respect for one another as learners with different styles, personalities, and viewpoints.

The SERVE-Leads Standards of Classroom Practice, the Principles of Learning (see text box), the characteristics of "authentic pedagogy" (text box), and Schlechty's WOW Framework described earlier are all attempts to define a powerful vision of effective classroom practice. Just as we talk about the need for a school vision to frame school improvement planning, teachers also need a classroom vision to focus ongoing dialogue about needed improvements.

Newmann and Wehlage's Authentic Pedagogy

A five-year study of classrooms identified dimensions of good teaching related to higher levels of achievement, especially in low-income or high-minority schools. According to the researchers, "authentic" pedagogy involves (a) engaging students in higher-order thinking, (b) addressing central ideas thoroughly in order to help students acquire deep knowledge, (c) fostering substantive conversation among students, and (d) connecting student learning to the world beyond the classroom. For more information, see *Successful School Restructuring: A Report to the Public and Educators* by Newmann and Wehlage, available at *www.wcer.wisc.edu/archives/completed/cors/default.htm*.

Team members collaborate to define a quality classroom.

Institute for Learning's Principles of Learning

The Principles of Learning are condensed theoretical statements summarizing decades of learning research. These principles are designed to help educators analyze the quality of instruction and opportunities for learning that they offer to students.

1. Academic Rigor in a Thinking Curriculum
2. Accountable Talk℠
3. Clear Expectations
4. Fair and Credible Evaluations
5. Learning as Apprenticeship
6. Organizing for Effort
7. Recognition of Accomplishment
8. Socializing Intelligence
9. Self-Management of Learning

For more information and an explanation of these Principles of Learning, visit www.instituteforlearning.org/pol3.html.

Reports from SERVE-Leads Districts

It would be difficult, even overwhelming, for a school or district to begin working on all their identified Standards of Classroom Practice at one time. The writings of Schlechty and others point to the importance of our second Standard of Classroom Practice: that work assigned to students should be purposeful, cognitively challenging, and result in useful feedback to students on how to improve. Thus, this Standard of Classroom Practice was selected as an initial focus. One of the tools SERVE located to help teachers reflect on the quality of their assignments is a process for collecting "typical" teacher assignments and scoring them on a number of dimensions identified in the research literature as being important for student achievement. (See *Taking a Close Look at the Quality of Teachers' Assignments and Student Work* on page 26.)

The following are reports from some of the districts participating in SERVE-Leads that are in the early stages of introducing initiatives to help teachers apply definitions of quality to the work they assign students. It is interesting to note that no districts took the same approach in beginning these conversations about quality with their teachers and schools.

In Mississippi, Petal School District Asks Teachers to Submit a Quality Assignment

Collect teacher assignments and offer teachers feedback from key district leaders as a first step.

Because it is a small district, Petal School District can move quickly. In the last several years, district leaders have introduced school leaders to the idea of "working on the work" through group study of Dr. Schlechty's book. Even though the district is a successful district (by state standards anyway—receiving the highest state accreditation rating of 5), leaders realize that in terms of getting all students to the level of functioning that will bring them success in a postsecondary setting, the district has much work to do. Group book studies and other strategies have been used to help school administrators and teachers realize the need to continuously improve.

While participating in a SERVE-Leads work session presented by Lindsay Clare of CRESST on using CRESST's rubric to score samples of "typical" teacher assignments in reading comprehension and writing, Ione Bond (Assistant Superintendent) and Tricia Bridges (Secondary Instructional Specialist) saw a tool to help implement their district's strategic plan. They quickly went to work. They wanted first to determine their teachers' perceptions of what quality work meant and second to gather baseline data in order to measure subsequent teacher growth.

In the spring of 2001, Dr. James Hutto (the superintendent), Bond, and Bridges asked all teachers to provide a list of attributes that characterize quality work. They decided that they would modify the CRESST process and ask teachers to submit a "quality" assignment instead of a "typical" assignment. Then they randomly selected teachers from each grade level and/or content area and asked them to submit an example of a quality assignment from their teaching with accompanying student work. Seventy-six percent of the teachers in the sample turned in assignments (with cover sheets describing the assignment).

Bond and Bridges then scored the assignments submitted using a rubric that a SERVE-Leads group had adapted from the CRESST rubric (see page 29). The rubric they used had seven dimensions—level of thinking required by the assignment, clarity of instructional goals and match to actual task demands, alignment with state/district standards, appropriateness of task demands for age, value beyond school, quality of assessment process, and overall learning value of assignment.

The average rating of all assignments submitted was a 2.47 (on a 3-point scale with 3 designating the highest quality). Considering that they used a 3-point scale rather than a 4-point scale, as used in the CRESST research, and that they asked teachers for a "quality" rather than a "typical" assignment, the results are not inconsistent with the CRESST findings in which higher-achieving schools had assignments averaging a 2.2 on a 4-point scale of quality.

Dr. Hutto, along with Bond and Bridges, then met with the teachers from each school who submitted assignments and gave them qualitative feedback on the strengths and weaknesses of the assignments submitted. Overall, the feedback given to teachers about their assignments was positive because many had submitted product-focused assignments, such as oral presentations and writing assignments that were cross-curricular in nature. One of the positive aspects of this initial effort has been to encourage all teachers to become familiar with the Quality Assignment Rubric used to score their assignments as a guide for future planning and development of assignments.

The feedback shared with the teachers included the following:

- At the lower elementary level, there were few examples of authentic assessments submitted and a lack of clear criteria for assessing the work students produced.
- At the upper elementary level, the review showed that some teachers were giving credit for just attempting the assignment, not holding students accountable and not applying the assessment criteria.
- The greatest weakness at the middle school was not setting high expectations for students, and, in general, the quality of the assignments was weak.
- The high school did not participate well in this district endeavor, which was disappointing; however, it was evident that the vocational teachers and Special Education teachers were doing a good job with quality work.

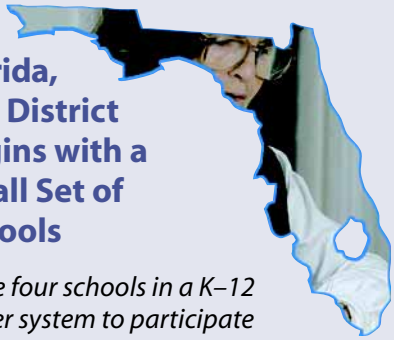
Having successfully introduced the idea of providing constructive feedback to teachers on the quality of assignments they give students, the next step was to find ways that teachers could begin to offer each other feedback in "safe," collegial settings. At a SERVE-Leads meeting in the fall of 2001, protocols developed by the Annenberg Institute for School Reform (*www.annenberginstitute.org*) for assisting teachers in structuring group feedback meetings were provided.

Beginning in late fall of 2001, teachers volunteered to begin working in small groups to test "protocols" for looking at student work. To date, only a few volunteers have shared teacher assignments and student work with their peers. The response from peer groups that have participated has been very positive. One teacher said: "By viewing actual work samples, I was able to connect the definition of quality work to the rubric. The words [in the rubric] came to life in the form of authentic student work." The volunteers who presented their assignments and samples of student work affirmed the value of this as a learning experience. One teacher presenter said: "Presenting this activity in front of my peers validated my understanding of quality work."

According to the Petal district leaders, this is still a work in progress.

(continued)

In Florida, Bay District Begins with a Small Set of Schools



Invite four schools in a K–12 feeder system to participate in a district research and development project.

Bay District Schools in Panama City, Florida, is a mid-size district serving a tourist-based community. It is a district with a history of site-based management. Because of this context, district leaders have a history of initiating district-wide initiatives as research and development efforts in a few schools. If the initiative works in those schools, it is offered to more schools. Because of its long history as a co-developer with SERVE on professional development for teachers in classroom assessment, we decided to continue this collaboration by developing and piloting tools and processes for helping schools address the Standards of Classroom Practice.

The plan for the spring, summer, and fall of 2002 is to work with school leadership teams to collect individual school baseline data and outline a unique approach for each school to begin "working on the work."

The leadership for this initiative includes newly elected Superintendent James McCalister, Director of Instruction Lendy Willis, and central office Resource Teacher Patricia Schenck, who has in the past headed up the district-wide training in classroom assessment and the development of district guidelines for good classroom assessment. In the fall of 2001, teams from the four pilot schools met with district leaders and SERVE staff to learn about the project. They subsequently explained the project to their respective faculties. SERVE and Bay County developed a process for collecting baseline data for each school on the Standards of Classroom Practice. These visits took place in February 2002. Reports back to the schools in May 2002 focused on understanding and use of state standards, strength of school and classroom vision, teacher working environment in terms of quality of opportunities to work together, and student learning environment, including quality of work they are asked to do.

Conducting a comprehensive review of classroom quality by using outside observers and teacher artifacts (assignment ratings) will help ground the next phase of professional development planning. The rationale for involving leadership teams from a K–12 feeder system of four schools is that students ultimately are educated across 13 years and many teachers, and there should be continuity in the quality of their opportunities to learn as they move from one level of schooling to the next.

In North Carolina, Lee County Public Schools Begins with Mentors

Start with mentor/mentee pairs of teachers.

Lee County is a mid-sized district in North Carolina with a history of focusing on teacher development. It was one of the first districts in North Carolina to implement a voluntary formative teacher evaluation component in all its schools. The district has done extensive work with the community and strategic planning, and it sees this initiative to improve the quality of work assigned students as supportive of its strategic plan. As with many districts, it has many ongoing initiatives that place demands on teachers, so the leaders decided to start experimenting with the idea of quality work within an existing program. The district's mentor program was selected. The mentor/mentee program in Lee County provides beginning teachers with non-evaluative and collaborative assistance from an experienced teacher.

District staff participating in a SERVE-Leads Consortium selected a group of ten mentor teachers and ten mentees (beginning teachers) in grades 3–8 who would focus on the quality of reading assignments. By encouraging beginning teachers (who will continue to enter the district's schools in high numbers) to think about the quality of work they give students, the district leaders believe that they can slowly develop a culture that includes critical collaborative examination of the quality of instruction and assessment.

During the fall and winter of 2001, the ten mentor teachers reviewed and adopted the SERVE-Leads rubric for evaluating teacher work products. During this review, they recognized the need to revisit their own beliefs about what constitutes quality teacher work and the need to include the concept of quality work as part of the mentor/mentee cycle of assistance. They agreed to solicit teacher work from their colleagues and collected about 50 samples from these volunteers. Samples of teacher assignments are submitted with a cover sheet completed by the teacher explaining the context for the assignment. This cover sheet (adapted from the CRESST process) has been developed as an electronic file at the request of the participating mentors in an effort to simplify the gathering of samples from teachers.

The group used these samples to practice applying the "quality" rubric. They will continue to practice with the rubric to enhance their ability to provide effective critiques. As part of the learning process, the group is looking at inter-rater reliability. After several work sessions, the mentors are beginning to feel comfortable with their ability to apply the rubric effectively. Their application of the rubric to the samples uncovered a wide range of quality. Of particular interest is the ongoing group discussion centered on an apparent lack of cognitive challenge in the work submitted.

As Lee County begins planning for the next year, the original project group of mentors/mentees will be expanded. Discussion will continue on how using this rubric fits into the reflective process of the mentor/mentee relationship. Lee County's quality teacher work initiative is an attempt to train and "retool" veteran, experienced teachers and to include the concept of quality work for students as part of the teacher induction process.

In Alabama, Mobile County Public School System Starts with a Small Core Group of Teacher Leaders

Start informally with a small group of volunteer teacher leaders.

Mobile County Public School System is a very large urban school district located in Mobile, Alabama. The district is undergoing severe budget cuts from the state and has very little money to spend on professional development, so the district had to start very small in its effort to get teachers involved in looking at the quality of work given to students. Two central office staff, along with a volunteer group of six teachers from the middle and high school, have sought out support from district personnel and made arrangements to work together to seek out methods to improve the quality of assignments teachers give students. They seek to heighten awareness and generate support for the need to improve the quality of classroom practices within their school district. The small group of teachers and central office staff meets regularly to bring work samples to the table and to try to apply rubrics, like that of CRESST, to improve the quality. They have discovered that they need to see more examples of high-quality assignments in order to get a better sense of where their assignments fall short, particularly in terms of "cognitive challenge." In other words, they have realized they need to see and discuss "exemplar" assignments.

In Georgia, Forsyth County Schools Will Involve Teachers in Examining the Quality of Standards-Based Units

First, build teacher capacity to develop standards-based units and then engage teachers in examining the quality of the units.

Forsyth County School District is located in Cumming, Georgia, in the Atlanta-metropolitan area. Currently a district of 18,000-plus, it has doubled in size in the past five years.

This district is one of the original ten pilot districts in the country involved with Phillip Schlechty's Center for Leadership in School Reform. The key premise of the reform is that the core business of schools is to provide students with quality knowledge work—work that is engaging and content appropriate. All improvement efforts in the district have reflected this premise. For example, the district has worked on:

- Identifying standards and benchmarks for student performance and collecting assessment data relative to student performance on these standards.
- Implementing personnel evaluation processes (the Professional and Leadership Appraisal Cycles) that reinforce the key premise.
- Developing an effective School Improvement Model.

In the 2001–2002 school year, a new system-wide staff development initiative was initiated under the guidance of Curriculum Coordinator Beth Reynolds. In the summer of 2001, each of the 21 district schools were invited to select four-to-six teacher leadership teams (called Core Teams) to participate in a yearlong, grant-funded, district-planned professional development program. To date, 115 teachers from the 21 schools have participated in five days of customized staff development. The Core Team strategy provides extensive and ongoing professional development on (a) building quality standards-based units, (b) designing quality standards-based assessments, and (c) aligning classroom assessment, grading, and reporting practices to standards.

These Year One Core Team participants will be provided two additional days of staff development during the summer of 2002 (for a total of seven days for the year) and will be provided district-level support throughout the next school year (Year Two) in the continued development and implementation of standards-based units and assessments. The district will engage a new cadre of Core Team participants in the fall of 2002.

Building on this comprehensive district-wide initiative to improve teacher capacity in implementing standards-based units, Reynolds adapted the Quality Assignment Rubric for use by Forsyth teachers in assessing the quality of their unit designs. Indicators built into the resulting Quality Unit Rubric include the principles of standards-based unit design as well as Schlechty's WOW Design Qualities. The Quality Unit Rubric, which will be piloted with Core Team participants as they complete their units, is structured to answer the following questions:

- Does the unit engage students in substantive content and provide opportunities to apply highly complex thinking?
- Does the unit exhibit alignment among learning goals, standards, assessment strategies, grading criteria, and instructional strategies?
- Does the unit exemplify a quality assessment environment that will lead students to experience success?
- Does the unit help students to apply and use knowledge beyond the school setting?
- Does the design of the unit support opportunities that engage and motivate students to learn?
- Does the unit show evidence that effective strategies are employed to meet the needs of all students?

Final Thoughts

The district summaries above provide a glimpse of how several districts are going about drawing teachers' attention to the quality of classroom learning opportunities. Clearly, there is no single right way. We call our research and development work with these SERVE-Leads districts the Standards of Classroom Practice Project. Helping districts develop such standards, assess the degree to which identified standards (their vision for quality) are found in classrooms, and develop learning opportunities for teachers so that they can "work on the work" is a

(continued)

critical aspect of internal accountability. There are many other aspects of district functioning that contribute to high levels of internal accountability (the quality of processes to recruit, hire, induct, and evaluate teachers; develop and evaluate principals; provide assistance to low-performing schools; support standards-based curriculum and assessment; engage the community; etc.). However, defining a vision for quality in classrooms and engaging teachers in working toward this vision is increasingly being recognized as a missing component.

*For more information about SERVE's Standards of Classroom Practice Project, contact Nancy McMunn (*nmcmunn@serve.org*) or Wendy McColskey (*wmccolsk@serve.org*).*

References

Angelo, T. A., & Cross, P. K. (1993). *Classroom assessment techniques* (2nd edition). San Francisco: Jossey-Bass.

Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education, 80*(2), 145-163.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.

Eckstrom, R., & Villegas, A. M. (1991). Ability grouping in middle grade mathematics: Process and consequences. *Research in Middle Level Education, 15*(1), 1-20.

Elmore, R. E. (1996). Getting to scale with good educational practices. In S. H. Fuhrman & J. A. O'Day (Eds.), *Rewards and reform: Creating educational incentives that work*. San Francisco: Jossey-Bass.

Elmore, R. E., & Burney, D. (1998). Continuous improvement in Community District #2, New York City. Retrieved from *www.Irdc.pitt.edu/hplc/hplc.html*

Good, T. L., & Brophy, J. E. (1986). *Education psychology* (3rd edition). White Plains, NY: Longman.

Institute for Learning. (2001). *Principles of learning*. Retrieved from *www.instituteforlearning.org/pol3.html*

Lewis, A. (2001). Add it up: Using research to improve education for low-income and minority students. Poverty and Race Research Council. Retrieved from *info@prrac.org*

Newmann, F., & Wehlage, G. (1995). *Successful school restructuring: A report to the public and educators*. Madison, WI: Center on the Organization and Restructuring of Schools, University of Wisconsin.

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

Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school*. Boston: Houghton Mifflin.

Rosenbaum, J. E. (1976). *Making inequality: The hidden curriculum of high school tracking*. New York: Wiley.

Schlechty, P. C. (2001). *Shaking up the school house*. San Francisco: Jossey-Bass.

Spillane, J. (2000). District leaders perceptions of teacher learning. *Consortium for Policy Research in Education Occasional Paper Series*. Retrieved from *www.upenn.edu/gse/cpre*

Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis, 21*(1), 1-27.

Wagner, T., & Vander Ark, T. (2001). *Education Week*. Retrieved from *www.edweek.org/ew/ew_printstory.cfm?slug=30wagner.h20*

The 2002 SERVE Forum on School Improvement

Designs for Success

October 20–22, 2002

SERVE will host the **2002 Forum on School Improvement** at the **Birmingham Sheraton** in the historic downtown of **Birmingham, Alabama, October 20–22**. Join us for in-depth courses of study related to Assessment & Accountability, Early Childhood Education, Improvement of Math & Science Education, Literacy, School Reform, Teacher Leadership, Technology, and much more. For more information, please e-mail Rebecca Rhoden Ogletree (bogletre@serve.org) or Cynthia Robertson (croberts@serve.org), or call them at 800-352-6001. **Watch our website, www.serve.org/forum**, for updates as the event draws closer.

Taking a CLOSE LOOK at the QUALITY of TEACHERS' ASSIGNMENTS and STUDENT WORK

By Lindsay Clare Matsumura, CRESST/University of California, Los Angeles

Quality of instruction is the heart of the school reform process. Students need qualitatively different learning opportunities if they are to achieve the goals set out for them by many states and districts. Despite the importance of quality instruction, however, limitations in available methodologies have made it difficult to collect information on the quality of classroom practice on a broad scale. For this reason, the nature and quality of students' learning environments have existed as a "black box" in many large-scale evaluation designs.

The National Center for Research in Evaluation, Standards, and Student Testing (CRESST) at UCLA has been conducting research aimed at developing a method for investigating the quality of students' learning environments by developing criteria for rating teachers' assignments and student work (Aschbacher, 1999; Clare, 2000; Clare & Aschbacher, 2001; Clare, Valdés, Pascal, & Steinberg, 2001).

CRESST's research has focused on developing indicators of instructional quality that can help monitor and support diverse school reform initiatives as well as that have the potential to direct attention to dimensions of practice that are germane to student learning (Linn & Baker, 1998). Our work so far has centered on language arts instruction in third and seventh grade. This research involved the collection of a number of different language arts assignments from teachers over the past four years, including "typical" writing and reading comprehension assignments, as well as "challenging" assignments. For each assignment submitted, teachers completed a one-page cover sheet describing their learning goals and assessment criteria. Teachers also submitted four samples of student work for each assignment—two of which they considered to be of high quality and two of which they considered to be of medium quality.

CRESST Dimensions of Quality

The framework for defining the quality of classroom assignments was rooted in research focused on the most effective instructional practices for promoting improved student learning (Newmann, Lopez, & Bryk, 1998; Porter & Brophy, 1988; Resnick & Hall, 1998; Slavin & Madden, 1989) and established professional standards for teaching (Danielson, 1996). Raters used a 4-point scale (1=*poor* to 4=*excellent*) to rate the following six dimensions of quality for each assignment submitted:

Cognitive challenge of the task. This dimension describes the level of thinking required of students to complete the task. Specifically, this dimension describes the degree to which students have the opportunity to apply higher-order reasoning and engage with grade-appropriate academic content material. For example, an assignment given a high score for cognitive challenge might require students to synthesize ideas, analyze cause and effect, and/or analyze a problem and pose reasonable solutions using content-area knowledge (e.g., comparing themes from different books, etc.). An assignment given a low score on this dimension, in contrast, might only require students to recall very basic, factual information.

Clarity of the learning goals. This dimension describes how clearly a teacher articulates the specific skills, concepts, or content knowledge students are to gain from completing the assignment. The primary purpose of this dimension is to describe the degree to which an assignment could be considered a purposeful, goal-driven activity focused on student learning. An assignment given a high score on this dimension would have goals that were very clear, detailed, and specific, and it would be possible to assess whether or not students had achieved these goals.

Clarity of the grading criteria. The purpose for this dimension is to assess the quality of the grading criteria for the assignment in terms of their specificity and potential for helping students improve their performance. Raters consider how clearly each aspect of the grading criteria is defined and how much detail is provided for each of the criteria. An assignment given a high score for this dimension would have grading criteria that clearly detail the guidelines for success and provide a great deal of information to students about what they need to do to successfully complete the task.

Alignment of goals and task. This dimension focuses on the degree to which a teacher's stated learning goals are reflected in the design of the assignment's tasks. Specifically, this dimension attempts to capture how well the assignment appears to promote the achievement of the teacher's goals for student learning. An assignment given a high score on this dimension would involve tasks and goals that overlap completely.

Alignment of goals and grading criteria. This dimension is intended to describe the degree to which a teacher's grading criteria support the learning goals (i.e., the degree to which a teacher assesses students on the skills and concepts they are intended to learn through the completion of the assignment). Also considered in this rating is whether or not the grading criteria include extraneous dimensions that do

not support the learning goals, as well as the appropriateness of the criteria for supporting the learning goals.

Overall quality. This dimension is intended to provide a holistic rating of the quality of the assignment based on its level of cognitive challenge, the specificity and focus of the learning goals, the clarity of the grading criteria, the alignment of the learning goals and the assignment task, and the alignment of the learning goals and the grading criteria.

CRESST's Findings

Quality of classroom assignments. For the first few years of our research, we focused on elementary and middle schools that served primarily lower achieving students (as assessed by standardized achievement test scores) from low-income families. In general, we found that the majority of assignments we collected from teachers were considered to be of "basic" quality (i.e., were scored a 2 on a 4-point scale) across the different dimensions (Clare, 2000). In the third year of our study, we scaled up our work to include elementary schools that served more economically advantaged students who were relatively higher achieving. Results (see Table 1) indicated that students from the higher achieving schools received higher-rated assignments overall, though there were certainly exceptions to this pattern (Clare et al., 2001). In other words, some teachers from the higher-achieving schools provided their students with relatively low-quality assignments, and vice versa.

Reliability and stability of the assignment ratings. Findings so far have indicated that there is an acceptable level of agreement between raters and good internal consistency for the classroom assignment rating scales (Clare, 2000; Clare et al., 2001). Results also indicated that three to four teacher assignments rated by at least three raters appeared to yield a stable estimate of quality. In other words, the estimated variance components based on the teacher assignment ratings showed that most of the variation in rating was accounted for by differences across teachers and not by differences across raters or assignment type. There is some indication, however, that this may only be true when teachers submit assignments that they themselves created, rather than submitting assignments they created in tandem with assignments they obtained from commercial sources. We are investigating this issue further in our current work.

Relation of classroom assignments and observed instruction and student work. To investigate the validity of the classroom assignment rating scales, we looked at the relation of classroom assignment ratings, observed practice, and the quality of student work. Our results so far indicated that the quality of teachers' assignments was associated with the quality of their observed instruction across a range of learning environments. The quality of teachers' assignments also was associated with the quality of student work (Clare, 2000; Clare et al., 2001). In other words, teachers who received high ratings on their assignments tended to receive high ratings for their observed practice as well. The quality of student work also tended to be higher for those teachers who submitted highly rated assignments.

Use of the Classroom Assignment Rubric and Future Research

More recently, we have turned our attention to investigating the use of the Classroom Assignment Rubric as part of the Los Angeles Unified School District's new accountability system. This method was piloted during the 2000–2001 school year with teachers in grades 4, 8, and 10. Future research also will focus on the use of the method for supporting self-reflection on the part of teachers. To this end, we hope to collaborate with SERVE to investigate teachers' use of the classroom assignment rubric in school-based professional development settings. Specifically, we intend to investigate whether this framework for looking at classroom assignments helps teachers create higher-quality learning environments for students by helping them think about:

- The kinds of skills their assignments promote
- How well their assignments promote their learning goals
- How clear and informative their grading criteria are to students
- The match between their grading criteria and their learning goals

Over the long term, we hope that attention to quality dimensions of classroom assignments will result in greater success for students on increasingly more challenging tasks.

Table 1: Comparison of Assignment Ratings

	Lower achieving (N = 13) M (SD)	Higher achieving (N = 16) M (SD)	P value
Cognitive challenge of the task	1.64 (.44)	2.23 (.61)	0
Clarity of learning goals	1.92 (.50)	2.32 (.56)	0.007
Clarity of grading criteria	2.37 (1.01)	1.94 (.66)	0.07
Alignment of goals and task	1.83 (.49)	2.17 (.48)	0.013
Alignment of goals and grading criteria	1.81 (.59)	1.71 (.55)	0.52
Overall quality	1.71 (.43)	2.21 (.48)	0

NOTE: Items were scored on a 4-point scale (1 = poor, 4 = excellent).

[1] Reproduced from Clare et al. (2001).

References

Aschbacher, P. R. (1999). *Developing indicators of classroom practice to monitor and support school reform* (Center for the Study of Evaluation Technical Report #513). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd edition). New York: MacMillan.

Clare, L. (2000). *Using teachers' assignments as an indicator of classroom practice.* (Center for the Study of Evaluation Technical Report #532). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Clare, L., & Aschbacher, P. R. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment*, 7(1), 39–59.

Clare, L., Valdés, R., Pascal, J., & Steinberg, J. R. (2001). *Teachers' assignments as indicators of instructional quality in elementary schools* (Center for the Study of Evaluation Technical Report #545). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Linn, R. L., & Baker, E. L. (1998). School quality: Some missing pieces. *CRESSTline.* Los Angeles: University of California, CRESST.

Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago schools: A baseline report.* Chicago: Consortium on Chicago School Research.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the Institute for Research on Teaching. *Educational Leadership*, 45, 74–85.

Resnick, L. B., & Hall, M. W. (1998). Learning organizations for sustainable educational reform. *Daedalus*, 127(4), 89–117.

Slavin, R., & Madden, N. (1989). What works for students at risk: A research synthesis. *Educational Leadership*, 46, 4–13.

Rating a Reading Comprehension Assignment

How would you rate this third-grade reading comprehension assignment?

A teacher asked students to read the book, *Dr. De Soto*, by William Steig, and individually answer the following basic comprehension questions:

- Why did Dr. De Soto have a lot of patients?
- How did Dr. De Soto work on extra large animals?
- What was wrong with the fox?
- Why do you think the De Sotos decided to treat the fox?
- What does the word *woozy* mean?

The teacher's stated learning goals for the assignment were to work on comprehension and answering in complete sentences. The teacher's stated grading criteria were: *I checked for answers that were in complete sentences and whether or not students understood the question and wrote answers clearly. Credit was given to certain students with acceptable answers even if they didn't use complete sentences*.

The assignment received a 2 because the questions did not require any higher level or inferential thinking on the part of students, and students were required to write only 1–2 sentence responses. Additionally, the level of the book was below what would be expected in third grade. Compared to stated learning goals for other higher-quality assignments, the goals for this assignment were not very elaborate or specific with regard to what aspect or theme of the story the teacher most wanted the students to comprehend. Comprehension seemed to be about

(Rating Reading, continued next page)

Rating a Writing Assignment

How would you rate this seventh-grade writing assignment?

Students were asked to write a five-paragraph essay on their dreams for the future.

The teacher's stated learning goals were to teach students, step by step, how to write a five-paragraph essay and to demonstrate the creativity and fun involved in essay writing. The teacher's stated grading criteria were as follows: *content is well explained, writers focus on what needs to be talked about, and writing process is completely done*.

This assignment received a 2 because students were not required to engage with substantive content material (e.g., research a topic) or apply higher-level thinking skills by, for example, arguing a position or comparing and contrasting different concepts, characters, or experiences in their writing. The focus of the assignment appeared to be more on just writing to a prompt than writing to learn or to communicate. Keep in mind that teachers were asked to submit a "typical" assignment given to students, and that over time, having students repeatedly working on five-paragraph essays in response to artificial prompts does not represent a high level of cognitive challenge. The teacher's stated learning goals for the assignment appeared to focus more on the activity of producing the essay, rather than developing content knowledge, developing students' thinking skills, or deepening their understanding of specific concepts. The criteria were considered to be of basic quality since the teacher did not clarify for students how she would decide how well they had explained their dreams or focused on what needed "to be talked about."

(Rating Writing, continued next page)

(Rating Reading, continued)

just getting the facts of the story right. The teacher's grading criteria did not seem very clear compared to other higher-quality assignments with very elaborate criteria for what the teacher expected. These grading criteria provided little information to students with regard to what they would need to do to be successful at the task. Also, there is some misalignment between the teacher's stated learning goals, which emphasized writing in complete sentences, and the grading criteria, which allowed for answers that were not in complete sentences.

(Rating Writing, continued)

As such, these criteria provided little information to students about what they would need to do to successfully complete the task.

CRESST Quality Assignment Rubric

	4	3	2	1
COGNITIVE CHALLENGE	Task requires strongly complex thinking as an extensive, major focus of the task. Students also engage with substantive content material.	Task requires complex thinking. Students may also engage with substantive content material.	Task requires some moderately complex thinking. Students may also engage with some substantive content material.	Task does not require any degree of complex thinking and/or does not engage students with substantive content material.
FOCUS OF THE GOALS ON STUDENT LEARNING	Goals are very focused on student learning. Goals are very clear and explicit in terms of what students are to learn from the assignment. Additionally, all the goals are elaborated.	Goals are mostly focused on student learning. Goals are mostly clear and explicit in terms of what students are to learn from the assignment.	Goals are somewhat focused on student learning. Goals are somewhat clear and explicit in terms of what students are to learn from the assignment. Goals may be very broadly stated. Or there may be a combination of learning goals and activities.	Goals are not focused on student learning, and are not clear and explicit in terms of what students are to learn from the assignment. Or all goals may be stated as activities with no definable objective.
CLARITY OF THE GRADING CRITERIA	Teacher's grading criteria are very clear, explicit, and elaborated.	Teacher's grading criteria are mostly clear and explicit.	Teacher's grading criteria are somewhat clear and explicit.	Teacher does not specify a grading criteria, or it is not possible to determine the grading criteria from the teacher's documents.
ALIGNMENT OF LEARNING GOALS AND TASK	There is exact alignment between the teacher's stated learning goals for students and what the task requires students to do. The task fully supports the instructional goals. The task and goals overlap completely—neither one calls for something not included in the other.	There is good alignment between the teacher's stated learning goals and what the task requires students to do. The task supports the instructional goals.	There is only some alignment between the teacher's stated goals and what the task requires students to do. The task only somewhat supports the instructional goals. Or the goal may be so broadly stated that the task and goal are aligned only at a very general level.	There is very little or no alignment between the teacher's stated goals and what the task requires students to do. The task does not support the instructional goals.
ALIGNMENT OF LEARNING GOALS AND GRADING CRITERIA	Excellent quality in terms of level of cognitive challenge, clarity, and application of learning goals and grading criteria.	There is good alignment between the teacher's stated learning goals and the stated grading criteria.	There is only some alignment between the teacher's stated learning goals and the stated grading criteria.	There is very little or no alignment between the teacher's stated learning goals and the stated grading criteria.
OVERALL TASK QUALITY	There is exact alignment between the teacher's stated learning goals for students and the stated grading criteria.	Good quality in terms of level of cognitive challenge, clarity, and application of learning goals and grading criteria.	Limited quality in terms of level of cognitive challenge, clarity, and application of learning goals and grading criteria.	Poor quality in terms of level of cognitive challenge, clarity, and application of learning goals and grading criteria.

THE SOUTHERN STATES SEMINAR: A Regional Strategy for Success

By Rebecca Rhoden Ogletree, SERVE Senior Program Specialist

At the SERVE Southern States Seminar in Destin, Florida, January 10–11, 2002, approximately 50 mostly state-level administrators from the six SERVE states of Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina had the opportunity to interact with researchers and to discuss state efforts to improve schools. The goal was to create a regional response to school improvement efforts and state policy recommendations—a "Southern Strategy for Success." The drive to achieve higher standards is fundamental to a school reform and improvement strategy. The keynote presenters, Andrew and Hathia Hayes of the University of North Carolina-Wilmington and John Poteat of the North Carolina Public School Forum, discussed the challenges to achieving higher standards, and participants met in groups to discuss the ideas presented.

Andrew and Hathia Hayes kicked off the seminar by presenting the findings from their study of Comprehensive School Reform Demonstration implementation efforts in 70 North Carolina schools. One problem they identified was that structural barriers in state policy and in local policy and practices are slowing down reform in many cases. They found that many of the schools' activities appeared not to contribute to the overarching goal of school reform and actually took time away from reform activities.

John Poteat, despite being a staunch supporter of the standards movement, discussed the unintended consequences that occur when higher standards are the driving force behind reform. One such consequence is the mounting opposition to the testing movement, which, according to Poteat, is fueled by legitimate concerns such as flaws in testing programs, debates over passing standards, charges of teaching to the test, disappearance of non-tested curriculum, and pressure on teachers and students. As a result, Poteat said, "The wealth of testing data now available to the public has spawned its own unintended consequences: feeding a negative image of public schools, causing a growing dissatisfaction in our minority community, and threatening to fracture the traditional pro-public school lobby."

Challenges to Reform and Accountability

Seminar participants identified the following factors that undermine efforts to achieve higher standards and ultimately successful school reform and improvement:

Policymakers' and teachers' interpretations of accountability do not always coincide.

Teachers feel pressured to improve test scores rather than teach to content standards. Teachers indicated, in nearly

every interview with the Hayes team, "a concern for penalties they may incur due to low student test scores." In their research, the Hayes team found that "state testing and high-stakes accountability programs contributed to low levels of reform." Reform efforts should be contributing to rather than competing with states' accountability efforts. Comprehensive school reform was meant to assist schools in raising student achievement and, therefore, should assist schools in meeting the requirements of the state accountability system.

Recruitment and retention of teachers is difficult.

Participants reported that many low-performing schools continue to fill positions with teachers working outside their content areas, and others are placing permanent substitutes in classrooms until qualified teachers become available. The Hayes team reported, "In several of these schools, staffing was mainly with initially-licensed teachers or lateral-entry teachers who revealed little evidence that they had the capability to make changes through their own efforts. Many of these teachers were merely trying to make it through each day and week." To recruit and retain teachers, many school systems are changing the "one-size-fits-all" pay schedule. According to Poteat, more schools are offering signing bonuses, providing higher pay for hard-to-fill jobs, and giving financial incentives to teachers willing to work in low-performing schools. In the future, a regional policy needs to be established that will raise the level of knowledge and capacity of all teachers in the Southeast so that they are better able to meet state standards.

Local and district leaders lack experience in reform.

State representatives noted that low-performing schools often demonstrate a lack of clarity in school vision and a failure to strategically address the needs of schools, especially when school leaders see improvement efforts as a matter of compliance rather than an initiative to change. The Hayes report indicates that principal leadership is essential for successful school reform: "An important proportion of the CSR schools did not have leaders skilled in (or willing to engage in) planning, developing, and communicating new and different visions of their school, nor were they expert in making decisions dependent on valid, goal-related data." Their study also revealed "a clear link between success and the expertise, experiences, and engagement of the principal, several factors in the history of the school, and the strength of the local education agency support." State and district leaders need to find ways to provide the professional development the school and district leaders need and to support administrators in their efforts as reform leaders.

> "The best-intentioned movement in modern education—the drive for higher standards—has driven public education to the brink. Unintended consequences of the movement could change the face of American education."
>
> *—John Poteat of the North Carolina Public School Forum at the SERVE Southern States Seminar*

Resources for professional development are limited.

Unfortunately, many school staff members are not receiving the continuous, focused, and meaningful professional development they need. One participant complained, "There is a failure to align money with the intense program assistance expected in individual schools." Furthermore, the Hayes team reported that staff development—as workshops primarily intended to introduce staff to models or practices, team-building retreats, and conferences—lacked the necessary intensity, duration, and/or quality, especially considering the degree of change being attempted and the degree of development teachers need. Private and federal resources need to be obtained to provide educators the professional development they need to carry out the states' efforts to raise student achievement.

Southern States' Progress

Despite these challenges, states have accomplished much as a result of focusing their efforts on improving low-performing schools. They have gained a clearer understanding of the needs of low-performing schools and a greater awareness of how to use disaggregated test data to improve instruction. Furthermore, they have developed focused school improvement plans, obtained professional development focused on the specific needs of schools and staff members as identified through assessments, exposed teachers to varied instructional procedures, and used resources more effectively. In addition, they have learned the following lessons:

- Ongoing professional development is essential.
- On-site assistance is most effective.
- Flexibility is essential because there is no one best strategy for all schools.
- Strong leadership must be developed at all levels, especially the district level.
- Schools must build capacity for continuous improvement.

Next Steps

In the last session, participants from the various states met in teams to address common concerns and to suggest approaches to address these issues regionally. Good ideas were generated, and participants agreed to continue to work toward successful strategies for regional success. To encourage the continued development of this regional vision, SERVE has established a website at **www.serve.org/LPS**. In addition, the SERVE policy analysts housed in each state department of education are partnering with SERVE School Development and Reform staff to plan the agenda for the next meeting. The policy analysts will ensure that participants in the January 2003 seminar will contribute to the creation of a shared regional vision.

SERVE SENIOR POLICY RESEARCH ANALYSTS ON REGIONAL ISSUES

Florida Approves New Grading Rule for School Accountability System

By Joel Overton, SERVE Senior Policy Research Analyst

In 1999, the Florida State Legislature passed the Bush/Brogan A+ Education Plan that established a school grading system in an effort to improve student learning. Under this plan, each school receives a grade of A–F based on its students' performance on the Florida Comprehensive Assessment Test (FCAT). The FCAT is the primary measure of the Sunshine State Standards, which were developed by the Florida Department of Education (in close consultation with teachers and other educators) to give parents, students, teachers, and administrators a clear understanding of what Florida's students should know at different stages of their academic careers. Student scores are classified into five achievement levels, with 1 being the lowest and 5 being the highest. On December 18, 2001, as part of the final phase of the Bush/Brogan A+ Plan, Education Commissioner Charlie Crist joined Governor Jeb Bush and fellow Cabinet members in approving changes to the state's school accountability grading system.

Under the new system, a greater emphasis is put on higher performance standards. Individual student progress is measured by learning gains, and a strong focus is placed on the lowest-performing students in each school. "This more diagnostic approach to measuring success in our schools will provide a better understanding of how well schools and students are performing," said Crist. "Accountability is the key, and this plan promises to open more doors for students as it will challenge them to work hard and achieve."

The Governor and Cabinet members who sit as the State Board of Education approved the changes to the school-grading rule—a major component of the Bush/Brogan A+ Plan for Education—by a unanimous vote. A school's grade will now be determined based on the following three criteria:

- The percentage of students meeting high standards on the FCAT in reading and mathematics in grades 3–10 and writing in grades 4, 8, and 10.
- The percentage of students demonstrating annual learning gains in FCAT math and reading in grades 3–10. Students demonstrate learning gains by advancing from one achievement level to the next, maintaining satisfactory achievement at level three or higher (on a scale of 1–5), or demonstrating more than one year's growth within achievement levels one or two.
- The progress in reading of the lowest 25% of students in each grade aggregated for each school. The minimum requirement for adequate progress is defined as at least 50% of these students making gains. If this requirement is not met, a school's advisory council must develop a component in its School Improvement Plan to achieve it. If a school is designated as a "B" or "C" school and does not meet this minimum progress for two years in a row, the school's final grade will be reduced by one letter grade.

"We have developed an exciting new concept in education that stresses both accountability and achievement, while not losing sight of the basic fundamentals of learning."

—Charlie Crist, Florida's Education Commissioner

The revision will provide benefits to students, parents, educators, and administrators. The new system measures the progress of students from year to year and clearly identifies students' strengths and weaknesses. This identification allows administrators to channel resources more effectively to facilitate and target improvement. "The changes we have made today benefit educators, students and parents alike," said Crist. "We can now measure the progress of each individual student and at the same time assess the performance of an entire school as well as place a special focus on those students who need the most help."

Individual student achievement, a key component of the new grading system, has long been an objective of the A+ Plan. Increased emphasis is placed on the achievement of the lowest-performing students in each school. This is clearly seen in the fact that in order for a school to earn an "A," it must meet the minimum requirement of at least 50% of its lowest performers making adequate progress. The difference between reading achievement among the lowest quartile and the reading achievement of the overall population of students tested must also be within ten percentage points of each other in order for a school to earn an "A." This focus on reading, which is a cornerstone of a solid education, aims to ensure that all children have the essential skills to succeed.

South Carolina Addresses African-American Student Achievement

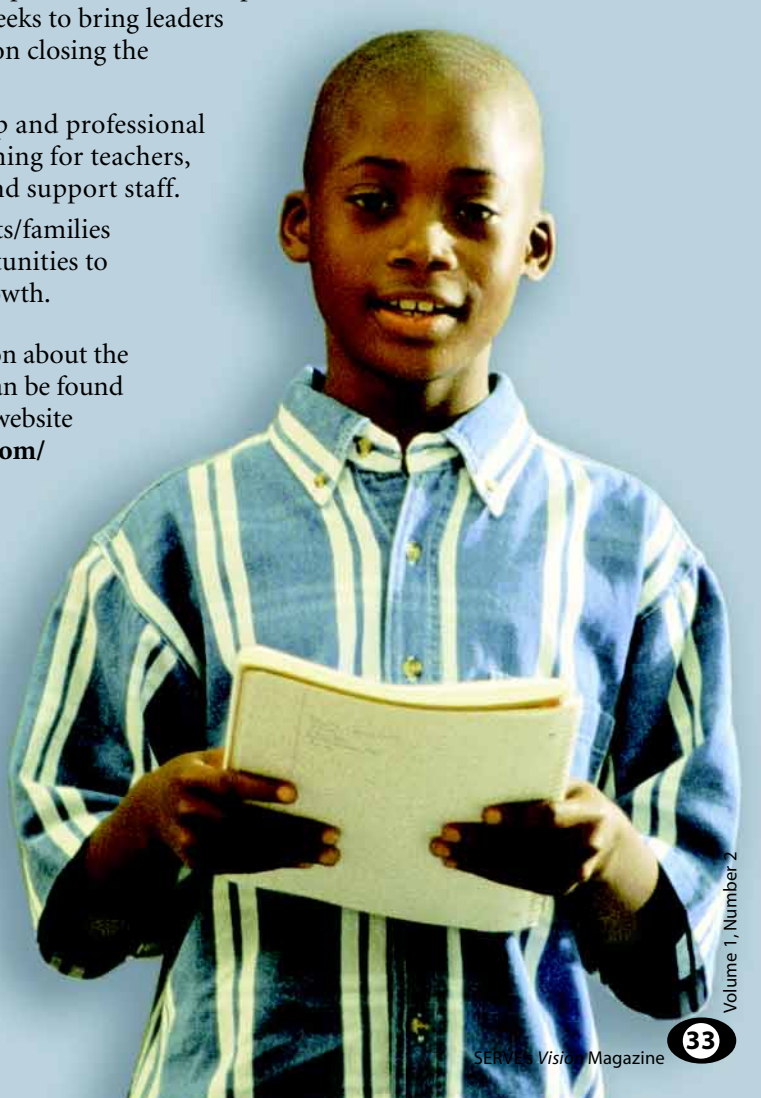
By Cindy McIntee, SERVE Senior Policy Research Analyst

In May of 1999, the South Carolina State Superintendent of Education, Inez Tenenbaum, established the African-American Student Achievement Committee to research, study, and develop strategies to close the academic achievement gap between minority and non-minority students in South Carolina. The Committee issued its report on May 30, 2001. The strategies, action steps, and recommendations issued in that report have been used to develop a framework for the work of the African-American Student Achievement Program (AASAP). The Program's advisory committee is comprised of educators from around the state, community leaders, faith leaders, South Carolina Department of Education professionals, and local and state agency representatives.

Some of the recommendations of the African American Student Achievement Program include:

- Collaborate with various groups to sponsor an annual conference on closing the achievement gap. The conference is planned for this year and seeks to bring leaders together to focus on closing the achievement gap.
- Provide leadership and professional development training for teachers, administrators, and support staff.
- Ensure that parents/families can receive opportunities to foster parental growth.

Additional information about the work of the AASAP can be found on the Department's website (**www.myscschools.com/offices/ssys/youth_services/aasap**).



New Testing Requirements

By Dr. Jeff Gagne, SERVE Senior Policy Research Analyst

The No Child Left Behind Act incorporates performance standards and consequences not only to Title I but also to all major programs. In addition to requiring tough corrective actions for chronically failing schools, it gives students in failing schools the right to either transfer to a better public school or obtain supplemental services.

The new federal law requires that states must implement annual reading and math assessments for grades 3–8. States will have until the 2005–2006 school year to develop and implement these assessments, and Congress has authorized $490 million for states to develop and administer them.

In addition to reading and math assessments in grades 3–8, states must also incorporate one other academic indicator. For secondary schools, it is graduation rates, and for elementary schools, it is an academic indicator determined by the state. In addition, states and school districts are free to incorporate additional indicators (including assessments) in the definition of adequate yearly progress. However, the only indicators that can put schools into school improvement, corrective action, or restructuring are the grades 3–8 reading and math assessments.

While states may select and design assessments of their choosing, these state assessments must be aligned with state academic standards; allow student achievement to be comparable from year to year; be of objective knowledge; be based on measurable, verifiable, and widely accepted professional assessment standards; and not evaluate or assess personal or family beliefs and attitudes.

Under the new law, states must define adequate yearly progress so that all students are expected to improve and that in 12 years all students will achieve at the state-defined "proficient" level on state reading and math academic assessments. Each state's definition of adequate yearly progress must apply specifically to disadvantaged students, as well as to the overall student population.

States set the starting point, or achievement "bar," to reach 100% proficiency and may choose where to set the initial bar based upon the lowest-achieving demographic subgroup or the lowest-achieving schools in the state, whichever is higher. States are free, however, to choose an even higher starting point. Once the initial bar is established, the state is required to "raise the bar" gradually and in equal increments to reach 100% proficiency. The initial bar must be raised after the first two years and then must be raised again, at least every three years.

In order to avoid "over-identification" of failing schools when students in a school are making significant academic progress, a "safe harbor" is allowed if students in the subgroups make a 10% reduction in the number of students not proficient. If students in a particular subgroup are 30% proficient and achieve a 7% increase in the number of proficient students, which is a 10% reduction in the number of students not proficient, they would be deemed to have made adequate yearly progress and would not be identified as failing. This provision has the added advantage of requiring larger gains for the subgroups farthest from proficiency while allowing for smaller gains for those closer to proficiency (where gains are harder to achieve).

In order to help turn around poor-performing schools, the new law also increases the current 0.5% set-aside of a state's total Title I allocation for school improvement activities to 2% for FY 2002–2003; it will increase to 4% for FY 2004–2007. In addition, the bill retains the separate authority for school improvement activities and authorizes it at $500 million in FY 2002 and such sums as may be necessary in FY 2003 through FY 2007. These funds will augment state and local efforts to provide technical assistance and improve schools identified as needing improvement. Technical assistance provided with these funds must be based on scientifically based research.

Requirements for Title I

- States must develop and adopt a single, comprehensive accountability system that will be used to hold districts and schools accountable for all students meeting state academic content standards.
- States must include in their standards specific goals for closing the academic achievement gap between whites and minorities and between rich and poor.
- Adequate Yearly Progress (AYP), the standard for judging educational success or failure, is redefined to ensure that every student reaches the state standard for proficiency within 12 years.
- Schools that do not meet the state standard for adequate progress but show at least a 10% improvement toward the goal of 100% proficiency for each subgroup ***and*** make progress on one other indicator will not be identified as in need of improvement.
- States must not only disaggregate their test data based on race and ethnicity, limited English proficient students, income, and disability, but also disaggregate the results for accountability. Each subgroup must meet the state standard for adequate progress for both mathematics and reading; otherwise, the school will be identified as in need of improvement.
- Content standards and assessments for math and reading in key grades

(elementary, middle, and high school) must be in place no later than the 2001–2002 school year.

- New annual state assessments in grades 3–8 in mathematics and reading must be developed and implemented by the 2005–2006 school year. Tests must measure students' progress in achieving proficiency on the state academic content standards.
- Science standards must be developed by the 2005–2006 school year.
- Science assessments in key grades must be developed and implemented by the 2007–2008 school year. Key grades indicate the administration of the science assessment at least once in elementary, middle, and high school.
- States must biennially participate in the National Assessment of Educational Progress (NAEP) in math and reading in the fourth and eighth grades to provide a benchmark for state assessments and standards.
- No less than 95% of all students must participate in the assessments, and the reasons for students not participating in assessments must be reported.
- States and local districts must compile and publish school-level report cards that contain meaningful and useful data about academic performance to parents and the public.
- An additional $500 million is provided to states to help turn around low-performing schools, and states must set aside up to 2% of their new Title I funds in FY 2002 (and up to 4% for the next four years) for the same purpose.
- Schools must meet rigorous school-improvement timelines. Schools will be given deadlines for implementing improvement plans and program reforms, and will receive additional technical and financial assistance if they initially fail to make adequate progress.
- Schools that fail for **two** consecutive years will give parents the option to transfer their child to a higher-performing public school with transportation costs provided.
- Schools that fail for **three** consecutive years will give parents the option to receive outside tutoring assistance from a state-approved list of providers that may include private organizations, non-profit organizations, and community-based organizations.
- Schools that fail for **four** consecutive years must undergo at least one corrective action, such as instituting a new curriculum, replacing the principal and some staff, or reopening the school as a charter school.
- Schools that fail for **five** consecutive years must continue the actions from the previous year and begin planning for restructure.
- Schools that fail for **six** consecutive years must be completely restructured, including instituting a new governance structure (such as a charter school or private management organization) and replacing all relevant staff.
- No later than four years after enactment (January 8, 2006), all teachers' aides must have (a) completed at least two years of study at an institution of higher education, (b) obtained an associate's or higher degree, or (c) met a rigorous standard of quality established at the state and local level that includes an assessment of math, reading, and writing. Paraprofessionals hired after January 8, 2002, must meet the new criteria.
- Up to 5% of any increase in Title I funding may be set aside by states to provide rewards to schools (and teachers in such schools) that substantially close the achievement gap between the lowest- and highest-performing students and that have made outstanding yearly progress for two consecutive years. In addition, 5% of any Title I funding increase may be set aside to award to successful teachers in schools that have been identified for improvement. Title II state activity funds may also be used to provide teacher awards.

Requirements for Title II

Teacher Quality

- States must set annual goals for increasing the number of highly qualified teachers and must have all teachers highly qualified by 2006.
- States must hold districts accountable for meeting teacher quality objectives and providing high-quality professional development to teachers and principals.
- Districts that fail to meet annual teacher quality goals for two years will be required to develop improvement strategies. Districts that continue to fail to meet their goals after three years will be forced to implement state-developed improvement strategies.

English Proficiency Requirements

- States must develop English proficiency standards and set annual achievement objectives for the development and attainment of English proficiency by limited English proficient (LEP) students. States must hold local districts accountable for annually assessing the English proficiency of all LEP students.
- Districts that fail to meet annual measurable objectives regarding English proficiency gains and academic achievement for LEP students after two years must develop improvement strategies. After three years of failure, districts must institute a new language instruction program and replace relevant staff.

Requirements for Other Titles

- Recipients of federal funds must set goals for performance in key programs, including Technology, Safe and Drug Free Schools, Reading, 21st Century Learning Centers, and Innovative Programs.
- States and districts that meet AYP goals will be rewarded by being given greater flexibility over the use of various funds.
- States that fail to meet AYP or their English proficiency goals shall be identified by the Secretary, expected to make improvements, and listed in an annual report to Congress.

Early Childhood Readiness Assessment Products

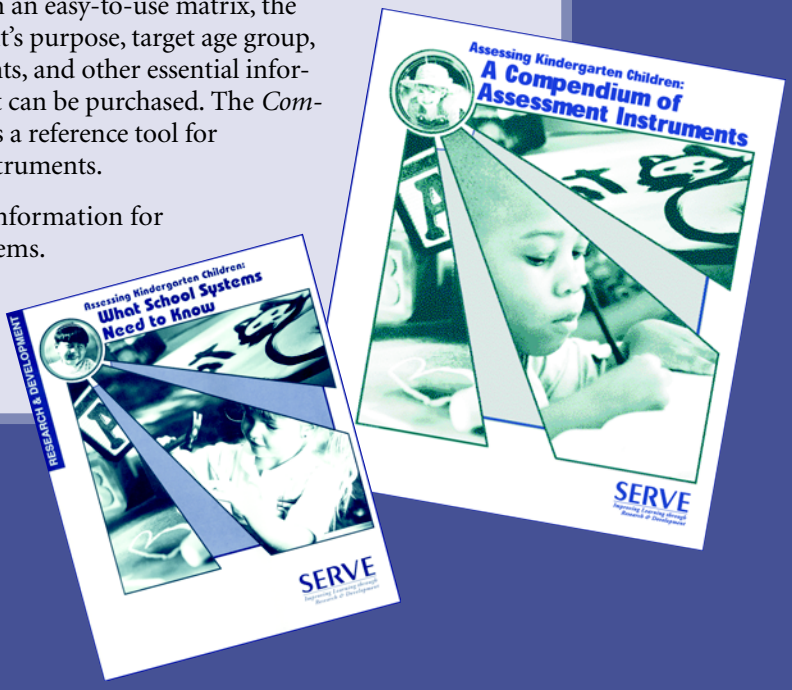
Schools and other early childhood programs across the region are struggling with questions about how best to assess young children and how to know if children are ready for success in school. Recognizing the critical need for valid and appropriate assessments, SERVE developed two publications to guide schools and other programs in selecting and implementing readiness assessment systems.

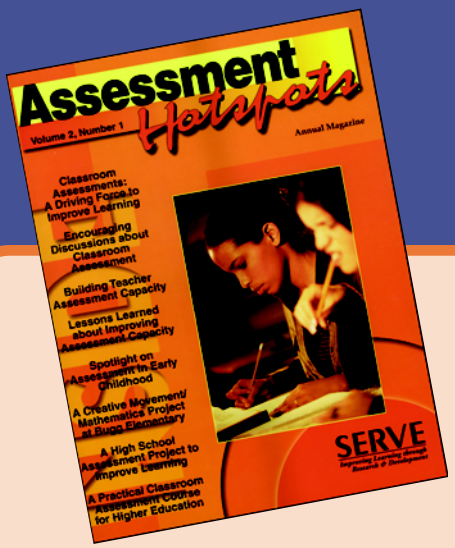
The first publication, entitled *Assessing Kindergarten Children: What School Systems Need to Know*, is a guidebook that presents principles for planning assessment systems and a practical process for designing assessment systems. The document combines information from a variety of sources to produce a practical, well-grounded planning guide. Sources of information for the guidebook include principles developed by the National Education Goal Panel Early Childhood Assessments Resource Group, numerous position statements on early childhood assessment, recent developments in early childhood research and theory, and SERVE's staff members' experiences working with states and local districts to develop assessment systems. *Assessing Kindergarten Children: What School Systems Need to Know* provides helpful suggestions for addressing complex issues related to early childhood assessment and includes a step-by-step guide for schools and programs to follow as they develop assessment systems.

Assessing Kindergarten Children: A Compendium of Assessment Instruments is a companion to the assessment guidebook. Selecting an appropriate assessment instrument is a critical part of designing a successful early childhood assessment system. However, there are a myriad of possible instruments and, prior to the publication of the *Compendium*, no centralized place for information on multiple instruments. The *Compendium* provides basic information on almost 40 assessment instruments that are commercially available. Arranged in an easy-to-use matrix, the *Compendium* presents information on each instrument's purpose, target age group, established reliability and validity, training requirements, and other essential information such as the cost of the instrument and where it can be purchased. The *Compendium* is designed to offer early childhood educators a reference tool for selecting the most suitable/appropriate assessment instruments.

Taken together, these publications provide essential information for programs in the process of designing assessment systems.

To request a print copy, please contact our Publications Unit at 800-352-6001 or visit us online at ***www.serve.org/publications****.*





Assessment HotSpots

SERVE has collaborated with schools, districts, state departments, and higher education institutions interested in improving assessment methods and practices at the classroom level. The primary purpose of this magazine, *Assessment Hotspots*, is to share the experiences of these educators. It contains eight articles describing reflections on initiatives by several schools and districts, one state department, and a teacher educator.

To request a copy, please contact our Publications Unit at 800-352-6001 or visit us online at ***www.serve.org/publications****.*

Competent Assessment of Reading: A Professional Development Module Training for Trainers

Learning to read effectively is an individual journey. Good feedback is an important part of that journey—especially for struggling middle school students trying to improve their reading comprehension skills. *Competent Assessment of Reading: A Professional Development Module for Middle School Teachers* helps teachers think about how to use assessment to improve students' reading skills. This new SERVE resource offers a continuous professional development program for fourth- through eighth-grade teachers in reading assessment. After an initial workshop ideally held at the beginning of the school year (for which all materials are included in the *Module*), teachers begin to implement new assessment practices. Focused attention is given to individualized reading assessment through "literature circles" and "individual reading conferences." This professional development module provides agendas and materials for follow-up sessions held throughout the school year to give support as needed, provide more in-depth explanations of assessment methods, review evidences, and continue professional growth in reading assessment.

The first Train the Trainer Session on this resource will be offered at the October 2002 SERVE Forum in Birmingham, Alabama. The one and a half day session is appropriate for any staff developer, administrator, teacher, or other educator who has a strong reading background, knowledge of classroom assessment, and who works with teachers on improving reading comprehension.

For more information, or to sign up for training, contact Nancy McMunn at 800-755-3277 or ***nmcmunn@serve.org****. This session is limited to the first 50 who sign up.*

The Professional Review Process: SERVE's Comprehensive System of Teacher Evaluation

Over the last ten years, SERVE has worked with educators in the Southeast to find ways to improve upon existing teacher evaluation systems. The result is *The Professional Review Process* for career teachers currently being implemented in districts in North Carolina and Mississippi.

What is different about SERVE's teacher evaluation model?

- Teachers use the same criteria (the Summative Scoring Matrix) for self-assessment and individual goal setting as principals use to provide feedback to teachers on their performance. The Summative Scoring Matrix provides teachers with guidance on concrete targets for improvement.
- SERVE's model of teacher evaluation includes both formative (feedback for individual professional growth) and summative (administrator judgment of teacher performance) components with the expectation that competent teachers will alternate formative years with a summative year so that principals are not evaluating every teacher every year.
- Teachers are actively engaged in examining their performance in areas like planning and assessment that require evidence beyond classroom observations. Having the principal engage the teacher in a final interview at the conclusion of a summative evaluation year creates an opportunity for professional dialogue and teacher reflection.
- The inclusion of Teacher Impact as a category in the Summative Scoring Matrix extends what is expected of teachers to how their teaching impacts students. The expectation that teachers should be able to report on and discuss knowledgeably their impact on student learning promotes a "bottom line" climate of continual learning and improvement in professional practices.
- Clear expectations for teacher performance are defined in 22 performance dimensions within six categories of teaching (i.e., Summative Scoring Matrix). Districts who choose to implement the SERVE model can modify the Summative Scoring Matrix to reflect their vision for good teaching. Individualizing the Summative Scoring Matrix to fit long-term district professional development goals allows accountability as well as feedback and support for teachers as they work toward implementing district-wide initiatives.

For more information, contact Dr. Barbara Howard at 800-755-3277 or ***bhoward@serve.org****.*

Using Research to Construct Quality Classroom Practices: Part I, Laying the Foundation

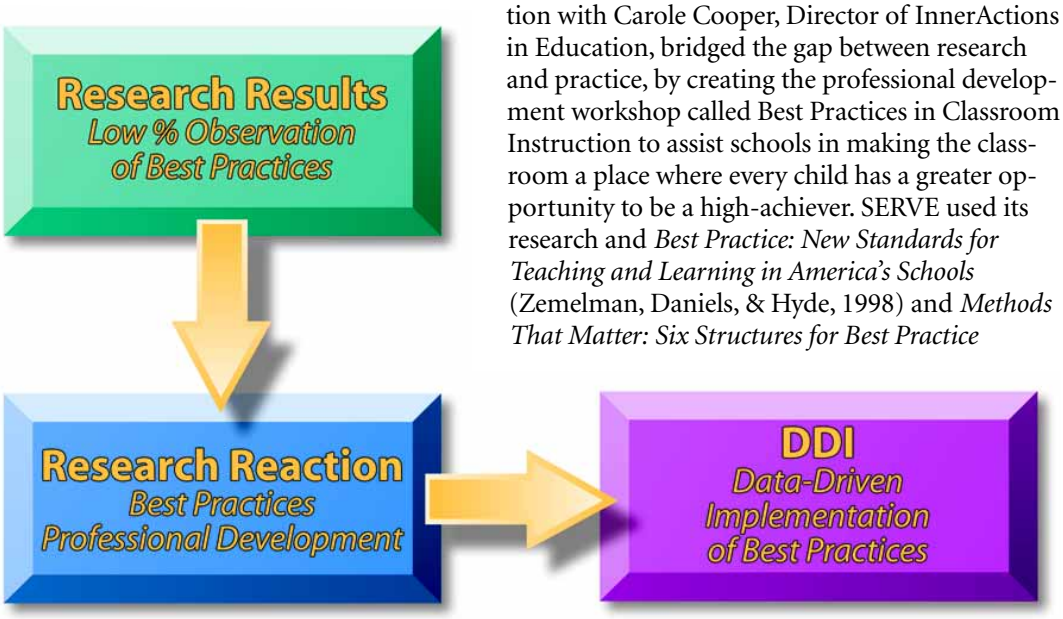
By Susan Martelli, SERVE Project Director; Scott Smith, SERVE Program Specialist; and Barbara Davis, SERVE Program Specialist

Recent research suggests that student learning is influenced most, not by family income or the level of a parent's education, but by good teaching. In other words, the quality of instruction in the classroom has the greatest influence on student achievement (Haycock, 2001). Late in 1997, the U.S. Congress passed legislation establishing the Comprehensive School Reform Demonstration (CSRD) program in an effort to enhance student learning. The nine original components of CSRD build on the effective school correlates (Brookover & Lazotte, 1979) and focus on the importance of rigorous curriculum and high standards for all students, effective leadership and efficient school management, ongoing and high-quality staff development, and sustained parental involvement. (For more information on these components, visit **www.serve.org/UCR**.)

From 1999 to 2001, SERVE conducted research on CSRD schools' progress toward implementing the nine components of school reform and the changes necessary to raise student achievement. Purposeful sampling was used to select 38 participating schools within our six-state region. Of the 38 schools, 18 had purchased school reform models from outside vendors, and five had developed their own models.

During the two years of research, trained and unbiased observers visited classrooms and completed the School Observation Measure© (SOM) (Ross, Smith, Alberg, & Lowther, 1999) and collected information on teachers' and professional staff's perceptions of CSRD program implementation at their school through the use of the Comprehensive School Reform Teacher Questionnaire© (CSRTQ) (Alberg & Ross, 1999). In addition, the School Climate Inventory© (SCI) (Butler & Alberg, 1991), which contains 49 items with seven dimensions, was also administered, and principal interviews and teacher focus groups were conducted. The information collected provided SERVE a combination of quantitative and qualitative data from each site.

As the research results unfolded, the following was noted: although nearly all of the CSRD programs being implemented at the school sites promoted a change to student-focused learning, the observations conducted in the classrooms indicated that little change had taken place in instruction. Across schools, the primary instructional orientation observed is best described as *didactic* and *teacher-led,* with independent seatwork the most prevalent student activity.



In response to these findings, SERVE in collaboration with Carole Cooper, Director of InnerActions in Education, bridged the gap between research and practice, by creating the professional development workshop called Best Practices in Classroom Instruction to assist schools in making the classroom a place where every child has a greater opportunity to be a high-achiever. SERVE used its research and *Best Practice: New Standards for Teaching and Learning in America's Schools* (Zemelman, Daniels, & Hyde, 1998) and *Methods That Matter: Six Structures for Best Practice*

> Phenomenal 'data-driven' workshop. Have been to AERA, NSTA, NSDC in Canada, CA, IN, FL, NC—none compare to this.
>
> *—2001 Forum Participant*
> *—Evaluation Report by Susan S. Harman, Ph.D.*

Classrooms (Daniels & Bizar, 1998) to design this workshop to give participants an increased understanding of classroom best practices; increased access to best practices theory, research, and materials; and increased understanding of the processes needed to support teachers' implementation of best practices. The Best Practices in Classroom Instruction workshop and materials have been delivered in a variety of formats from 2000 through 2002. Participants have consistently rated the workshops as highly informative, timely, and useful. In response to participants' feedback and research from the site visits indicating that school staff need to learn "how to use data to develop and continuously improve their school reform efforts" (CSRD in the Field, 2000), SERVE staff expanded the Best Practices workshop to include training in the Data-Driven Implementation (DDI) process.

Schools implementing the DDI process will analyze data from their schools, including student work. The data will guide the staff working in learning teams to create and carry out an action plan for implementing best practices in classroom instruction. SERVE staff's role in the implementation of DDI in each of the selected schools includes providing technical assistance via the Web; correspondence through electronic mail, telephone, and fax; and periodic visits that focus on the needs of each of the DDI sites. In addition, SERVE will continue to collect data via the SOM, SCI, principal interviews, and teacher focus groups and to provide an annual report to the schools each year that helps drive the decisions being made toward the goal of increasing student achievement.

SERVE staff members are rolling up their sleeves and planning to use Data-Driven Implementation to improve the quality of instruction and learning in the classroom.

While the number of schools touched by this first pilot program is small, the potential is infinite.

How Teaching Matters: Bringing the Classroom Back into Discussions of Teacher Quality published by Miliken Family Foundation and Educational Testing Service (ETS) suggests that high-quality professional development is the way to improve teaching, and that policymakers need to "stop scratching the surface of teaching and learning through superficial policies and instead roll up their sleeves and dig into the nature of teaching and learning by influencing what goes on in the classroom" (*Education Week*, October 2000).

THE VISION

SERVE
1203 Governor's Square Blvd.
Suite 400
Tallahassee, FL 32301

PRSRT STD
US POSTAGE
PAID
PERMIT HOLDER