From the Directors —
Effects of Performance Assessments On Teachers and Students
by Robert L. Linn and Eva L. Baker

Proponents of assessment reform and advocates for performance-based assessments as important components of broader educational change readily acknowledge the shortcomings of test-driven reforms in the past. Indeed, a now familiar litany of unintended negative side effects of high-stakes uses of standardized tests generally accompanies calls for moving to a performance-based assessment system.  WYTIWYG — what you test is what you get — is advanced both in the argument that standardized tests have negative side effects due to the resulting overemphasis on simple facts and discrete, low-level skills and in the argument in favor of a switch to assessments that are “worth teaching to.”

It was the concern about the effects of assessments on teachers and students that led to the heavy emphasis on consequences in our CRESST validation criteria.  A critical and still largely untested assumption in the movement toward performance-based assessments is that these assessments can have a positive impact on instruction and learning without introducing unintended negative side effects. Hence, a central strand of the CRESST research agenda is the accumulation of evidence that will seriously test this important assumption. The work in progress on the project summarized in this issue of CRESST Line illustrates a major component of this research (see page 2, Bumps and Bruises: Challenges to Implementing Performance Assessments in the Classroom).

Previous experience in the evaluation of educational programs underscores the need to attend to specifics of implementation at the local level. The black-box approach, where a program was introduced and then evaluated in terms of student outcome measures with little or no attention to what went on in between, proved to be sterile in the early days of the mandated evaluations of federal educational programs in the 1960s. The actual implementation of a

(continued on page 7)
Bumps and Bruises: Challenges to Implementing Performance Assessments in the Classroom

Ronald Dietel

"There were five areas of assessment with which every teacher struggled: what to assess, how to assess, how to score the assessment, how to keep track of the results and how to report results."
—Roberta J. Flexer & Eileen A. Gerstner

Change is not easy. Bombarded with overwhelming statistics about the dangers of smoking, the dangers of high-fat food, the dangers of not wearing a seatbelt or the importance of exercising, many Americans continue to smoke, continue to eat high-fat foods, refuse to wear seatbelts, and get much of their exercise by walking from their car to their office. So should we expect that creating classroom assessment change will be any easier? Probably not. The advice of Washington Irving in Tales of a Traveler may be invaluable as it relates to the change process: “There is a certain relief in change,” wrote Irving, “even though it be from bad to worse; as I have found in traveling in a stagecoach, that it is often a comfort to shift one’s position and be bruised in a new place.”

Based on evidence in a series of four new CRESST reports, teachers are suffering quite a few bumps along their own stagecoach journeys to implement performance assessment. Many feel overwhelmed with developing standards and performance tasks, asking themselves whether or not the results will be worth the bruises.

Headed by CRESST partner Lorrie Shepard, a team of University of Colorado at Boulder researchers have released a series of four new reports1 that may help answer the “worth” question.

Of important note, the researchers found that teachers in all three schools contended with many similar issues and problems.

“‘There were five areas of assessment with which every teacher struggled,’” wrote Flexer and Gerstner in their report Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics: “what to assess, how to assess it, how to score the assessment, how to keep track of the results, and how to report the results.”

Another key implementation issue was the conflict that teachers had with their own belief structure about assessment and with the expectations of other teachers and parents.

“The good news,” wrote Flexer and Gerstner, “is that we saw lots of change (albeit some in small steps)... The discouraging but expected non-news is that change is slow and nonlinear. And the more profound the change, the slower [the process].”

Lessons from these four reports will likely aid states, school districts and individual schools in the process of developing their own performance assessment programs.

1 Titles and prices of the reports are listed in the box to the left.
What to Assess

Relieved of any district requirements for standardized testing, teachers in the project were freer to decide how to teach the district curriculum and how to assess student learning. Teachers at the three schools responded differently. At Pine School, for example, teachers had difficulty accepting the new curriculum based on the NCTM standards and preferred teaching to the previous curriculum guide that emphasized more traditional mathematics. They tried to teach in the old way, adding on elements of the “new curriculum,” but eventually realized there was not enough time to do both.

Walnut School had a different problem. “Walnut teachers’ first dilemma arose,” wrote Flexer and Gerstner, “when they tried to establish an assessment framework...They worked diligently and probably too intensively on the framework, in that they divided the content dimension into smaller and smaller categories, as if they were defining each task, rather than dealing with a broad area of accomplishment.”

The result was a framework that didn’t reflect what the teachers taught or what students needed to know.

The result was a framework that didn’t reflect what the teachers taught or what students needed to know.

Bumps and bruises also arose over the use of different instructional strategies, oftentimes resolved through discussion and compromise. “I don’t want to do all problem solving every day,” said one teacher... “And I don’t want them to do all pages out of their books every day...I think they need a balance.”

Eventually, teachers compromised, dropping some topics from their standard curriculum so that there would be more time for problem solving.

The struggle with deciding what to assess is probably a necessary process for teachers implementing a new assessment system. Teachers learn to answer the what to assess question by working together, just as they would to resolve other instructional problems. Progress occurs through discussion and in some cases argument. But progress does happen. By the end of the first year of the project, one teacher answered the what to assess question by saying, “Well, we first have to ask, what do we want kids to know?”

How to assess

Knowing how to assess student learning has challenged teachers for years. Many teachers arrive in the classroom with little or no background in the evaluation or testing of children. Performance assessments have only complicated the process, primarily because an adequate supply of performance tasks is not readily available, and such tasks are arduous and time-consuming to develop.

In this study, the research team mistakenly assumed that most teachers would willingly develop their own tasks, according to Flexer and Gerstner. Although one teacher created some excellent tasks on her own and shared them with other teachers, most teachers did not have adequate time or training to create tasks that would measure the NCTM-like outcomes. But teachers did make effective use of assessments developed by others or methods recommended by the research team. For example, many teachers used “running records,” notes of mistakes made as the child reads, to assess student reading skills, especially for below grade-level children. Collected over time, running records showed a child’s reading progress or lack of progress.

Teachers also struggled with their beliefs about the relationship between instruction and assessment, oftentimes reverting to a traditional concept that students have to be prepared for the assessment tasks instead of viewing the tasks themselves as valuable learning experiences. “The teachers,” wrote Flexer and Gerstner “seemed to feel that children cannot figure things out if they haven’t been shown how to do them first, that they can’t hypothesize, conjecture, and test their hypothesis.”

But gradually teachers discovered that students were capable of complex problem solving, exemplified by one teacher’s observation of two children who made up a rule for finding prime numbers. “How interesting,” wrote Flexer and Gerstner, “that these children had stumbled on a problem that has concerned mathematicians for hundreds of years!”

How to score it

CREST research has continuously focused on the importance of developing specific, well-defined scoring criteria of student work matched to performance standards. But teachers still struggle, and will likely always struggle, with grading student work. In consort with the researchers involved in this study, teachers gained greater insight into the scoring process as represented by comments from one teacher who said: “The fact is that kids before they start need to know the [scoring] criteria. And lots of times we don’t do that.”

Teachers developed important links between scoring and standards.

Important links were also developed by teachers between scoring and standards. Hiebert and Davinroy in Dilemmas and Issues for Teachers Developing Performance Assessments for Literacy, wrote:

“Teachers raised the need for more guidance on scoring, and the final session of the semester was devoted to scoring of [student] sum-

(continued on page 4)
**Teacher Changes in Beliefs, Instruction, and Assessment**

One of the purposes of the CRESST/University of Colorado at Boulder study (feature article on page 2) was to document the results and changes that occurred to teachers and students as their schools implemented performance assessments into the classroom. In addition to the first two reports discussed in the feature article, the following information includes excerpts from a third study, CSE/CRESST Technical Report, 366, Teachers’ Ideas and Practices About Assessment and Instruction. The authors of this report are Hilda Borko, Maureen Flory, and Kate Cumbo, University of Colorado at Boulder.

### Changes in Beliefs

- The teachers...fell into two categories: (a) teachers grappling with ideas at deeper levels of beliefs and using them in their general movement toward teaching from a more constructivist perspective of learning; and (b) teachers who adopted the ideas into their current belief system, revising them to make them their own.

- The change for some Walnut School teachers was not so much in adopting different beliefs, but in moving further along the lines of the constructivist beliefs they already held. As they saw more positive effects in their students from new changes, they were reinforced by the changes and continued to do more and more.

### Teachers’ Comments Reflecting Change

- I think there have been times [in the past] that we have said, “Tell me how you got that,” and the kid explains that and we say, “Great.” This year, we say, “Tell me how you got that [on paper]” and I have proof to show parents and to show me.

- Today we solved a problem and we got six different explanations of how you could have possibly solved it. In my mind, math has been, in the past, right or wrong, and I’m really trying to encourage them [students] to think flexibly, to be flexible in their thinking that, well if it didn’t work this way I could try this, or if it worked this way could it work another way?

### Other Results

- All teachers developed criteria for summaries. They shared these criteria with their students who, to varying degrees, participated in the evaluation of their own written summaries. As one teacher at Walnut explained, students gained “an understanding of what teachers expected from them...I don’t think some kids ever had a clue before this” as to how we got their grades in reading.

- ...teachers frequently expressed surprise at how well the students were doing, how much they knew, and how much they enjoyed math. The teachers see evidence that the children are getting better at problem solving and at writing explanations for their solutions to the problems.

- Perhaps the most consistent change associated with participation in the CRESST/CU [University of Colorado at Boulder] project, in either reading or mathematics, was that all teachers were using running records (although to varying degrees) with their below-grade readers.

### Scoring performance assessments caused teachers to see students in a new light.

Consequently, the scoring of performance assessments caused teachers to see their students in a new light, realizing that many were not performing as well as expected. Perhaps more importantly, teachers increased their own level of expectations for students.

### Tracking the results

As noted by other recent CRESST reports, particularly the Vermont portfolio study,3 keeping track of the results...
of performance assessment work is onerous. In addition to the physical storage problems inherent to performance tasks, a greater problem has been knowing the best way to record and report the results of student work. It is probably safe to say that there is no single "best" method.

"The classroom comes first and extras come later and lately there have been lots of extras."

In the Colorado study, all teachers found record keeping to be a bumpy, bruising encounter. Part of the challenge was that teachers were being asked to change their patterns of record keeping from recording scores on worksheets, timed tests, and end-of-chapter tests to creating and maintaining detailed running records—summaries written by students of what they had read, or written explanations to tasks of problem solving and computing, and teachers notes on observations. In many cases teachers felt overwhelmed by the process, many of them simply adding on the new assessments to the old. As one frustrated teacher said: "The classroom comes first and extras come later and lately there have been lots of extras."

Researchers Hiebert and Davinroy responded to the record-keeping and time problem by devoting an entire workshop to showing teachers how to incorporate running records into everyday tasks. Afterwards, "Pine teachers returned to a focus on the task," wrote the researchers: "...the nature of changes in instructional practices was a prominent theme in this [the next workshop] session."

Showing teachers that performance tasks can be incorporated into instruction will not eliminate the burden of record keeping, but it may ease the problem considerably.

Reporting the results

Results from the performance tasks created difficult problems for teachers involved in the literature project. Ironically, the performance assessments were doing what they were supposed to do, revealing new understandings of student abilities. Unfortunately, in most cases, the assessments showed that students were not performing up to the standards expected by teachers. Hiebert and Davinroy wrote: "The assessments showed Spruce [School] teachers that some of their students were not able to read beyond a very rudimentary level. They wondered if they should share that information with parents or with the students."

The issue of reporting the results to parents or students was not resolved during the initial part of this study, but illustrates an unexpected dilemma that may arise when using performance assessments. An added problem was that the performance tasks made many students aware of their own shortcomings and may have bruised their self-images. Again, there is no easy solution to this problem, but future study will document how teachers deal with some of these consequences.

Conflict

Even though participating teachers from all three schools were volunteers, conflict was inevitable. "In the first six months of the project," wrote Flexer and Gerstner, "issues and dilemmas arose for this [Pine School] group of teachers in almost every category. Rarely in these discussions did all the teachers take the same side. The more experienced teachers were the more vocal in these discussions."

The research team reported that teachers had many disagreements about frameworks, standards, instruction, and different demands placed on them by different sources.

"Two things may have been happening," suggested Flexer and Gerstner, "at one of the schools. It was the teachers' perception that they were being pulled in a direction opposite to that of the other teachers in the school and of their students' parents, and the teachers may have been losing control of their instruction."

Much of the problem was summed up by one teacher: "It's like we're being geared to do problem solving with the kids and all that and then teachers in upper grades are upset because they're [students] coming into them and not having the computational skills that they think they should have...and then parents come in and say 'I don't know why my child doesn't bring home 25 addition problems every night to work on, what good is this going to do to have them count the legs on this animal?'"

"These teachers," explained Flexer and Gerstner, "were finding it very difficult to reconcile other teachers' expectations, parental demands, school goals, and their own beliefs with the ideas being presented by the researchers and the mathematics specialist. The teachers at this school were having a difficult time within the culture of their school when nominally it appeared that they had the district's and school's support, but in practice, they felt subjected to countervailing values."

This dilemma also remains unresolved, but suggests an expected consequence of teachers working to implement new assessment programs.

Change is occurring; we just don't know if it's a drop, a stream, or a flood.

Results

The bruises experienced by teachers involved in this research study were considerable. But were the results worth the bumpy journey? The researchers found significant evidence to indicate that at least some teachers were changing not only their beliefs about assessment but their instructional practices as well. As Flexer and Gerstner wrote: "If change is measured by how well the ideas about performance assess-
ment become integrated into the teachers’ repertoires, then we cannot look at schools but must think about individual teachers. Each school had a teacher or two or three who appeared to be engaged by performance assessment.”

Whether or not these changes represent systemic changes in teachers’ beliefs and assessment patterns should become more apparent as the study continues. Change is occurring; we just don’t know if it’s a drop, a stream, or a flood.

**Parent Opinions About Assessment**

In this fourth study from the series on the effects of performance assessment, CRESST/University of Colorado at Boulder researchers Lorrie Shepard and Carribeth Bliem set out to ascertain parents’ opinions about assessment, including their opinions about standardized tests versus performance assessments. Parents were given an opportunity to review performance assessment tasks in math and reading and decide what type of assessment, standardized or performance, was most suitable for classroom and district uses. Following are some of the results of the study abstracted from CSE/CRESST Technical Report 367, Parent Opinions About Standardized Tests, Teacher’s Information and Performance Assessments.

- Third-grade parents overwhelmingly believe that they learn the most about their child’s progress by listening to the teacher; 77% rated this source of information as very useful and 94% rated hearing from the teacher in the top two categories. All three informal sources of information about student progress—report cards, talking to the teacher, and work samples—received strikingly higher usefulness ratings than did standardized tests.

- By far, the majority of respondents preferred performance assessments. Even respondents who preferred standardized tests noted that performance assessments would stimulate children’s imagination or make them think.

- When parents were given the opportunity to study performance assessments, they offered sophisticated insights. They cited performance assessments’ ability to provide diagnostic evidence about a child’s thinking but also noted the potential for unfairness when having to write about math.

- A small subgroup of parents preferred the use of standardized tests for both district and instructional purposes because standardized tests are more cut and dried, more aligned to instruction, and easier. A few parents believed that standardized tests were better for math because “there is only one right answer in math.”

- **A key lesson** These three schools had the advantage of working with a research team experienced in the content areas and in working with teachers to create assessment change. Additionally, teachers and their principals had volunteered to be part of the study. Other schools, districts, and states attempting to implement performance assessments into the classroom may not be so fortunate. Research indicates that teachers need professional, long-term assistance to implement change. And lots of it.

“**Change does not roll in on the wheels of inevitability, but comes through continuous struggle.”**

“Change does not roll in on the wheels of inevitability,” said Martin Luther King, Jr., “but comes through continuous struggle.” Anyone desirous of implementing performance assessments into the classroom will likely face a similar challenge.

**CRESST Internet Server Usage Grows**

Newly compiled usage statistics indicate that the CRESST Internet gopher server has expanded to meet the public’s assessment needs. Average connections to the CRESST server now top over 140 people per day, suggesting that electronic communications has evolved into a critical medium for disseminating the results of CRESST research.

Among the most popular files on the server is the “Alternative Assessments in Practice Database” with listings of alternative assessments from over 250 sources. A recently added file, “What Is New at CSE/CRESST,” enables frequent CRESST gopher users to quickly access the latest in CRESST research and reports.

Computer users with Internet access and a gopher client program may access the CRESST server by starting a new gopher and pointing at: gopher.cse.ucla.edu.

CRESST will continue to update the server and has recently begun development of a World Wide Web server that incorporates audio and visual technology over the Internet. Stay tuned for future developments.
New Report on Equity and Assessment

Assessment, Equity, and Diversity in Reforming America's Schools
Linda F. Winfield & Michael D. Woodard
CSE/ CRESST Technical Report #372, ($4.00)

A national standards and assessments recently proposed as a strategy for improving schools in the United States have been accompanied by considerable tension between the goals of educational quality and equality of opportunity. "Proposed federal policies for implementation [of new standards and assessments] raise serious concerns about the extent to which national standards and assessments alone will help improve the quality of public education for all," write CRESST researchers Linda Winfield and Michael Woodard in their new report, Assessment, Equity, and Diversity in Reforming America’s Schools. The authors question whether or not some elements of the Goals 2000 program may "serve to deepen the already severe educational and economic cleavages that exist in this nation, especially along racial/ethnic lines."

Providing a framework to review equity, diversity and assessment, the authors present a variety of research findings to support their position. Findings of a national study of promising programs in disadvantaged urban and rural schools, for example, suggest that opportunity to learn is influenced by factors such as level of implementation, budgets, staff development, and administrative support. Winfield and Woodard believe that by omitting these factors from consideration in reform measures such as the Goals 2000 program, existing inequalities will be further exacerbated by creating additional barriers and limiting upward mobility opportunities for minority students.

Rather than pursue national standards and assessments, the authors suggest that reformers focus on policies and practices that have a greater probability of improving school learning and achievement, including equitable school financing, improved funding for curriculum development, and increased staff development for both teachers and administrators in content area assessments.

The authors conclude: "Only when policy makers consider opportunity-to-learn standards as important as implementing national standards and assessment, will we ensure that those students and individuals historically disenfranchised will share in the American dream of opportunity for educational achievement and economic success."

To order this report, or for a complete list of more than 140 technical reports, monographs, and resource papers, please call Kim Hurst at (310) 206-1532, or write to CSE/ CRESST Annex, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, CA 90024-4108.
National Testing Standards Under Revision

The American Psychological Association (APA) announced that a new committee recently met to update the Standards for Educational and Psychological Testing.1 CRESST Co-director Eva L. Baker is serving as co-chair of a joint committee to update the test standards together with co-chair and APA Past President Charles Spielberger, University of South Florida.

Sponsored by the APA, the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME), the Joint Committee on the Standards for Educational and Psychological Testing is addressing a range of issues including the uses of tests for decision making and the application of the standards to public policy changes. AERA, APA, and NCME represent over 140,000 researchers, practitioners, educators, consultants and students.

“Our challenge,” said Dr. Baker, “is to develop an approach to the revision of the Standards that is inclusive and addresses concerns of technical and user groups. At the same time, we want to create a process that will result in timely completion of a high-quality product,” she added.

Additional members of the board include a host of distinguished measurement experts including: Albert E. Beaton, Boston College; Lloyd Bond, University of North Carolina-Greensboro; Leonard S. Feldt, University of Iowa; David S. Goh, Queens College, City University of New York; Bert F. Green, Jr., the Johns Hopkins University; Edward H. Haertel, Stanford University; Jo-Ida Hansen, University of Minnesota; Sharon Y. Johnson, Detroit Public Schools; Suzanne Lane, University of Pittsburgh; Manfred J. Meier, University of Minnesota; Pamela A. Moss, University of Michigan; Esteban L. Olmedo, California School of Professional Psychology; Diana Pullin, Boston College; and Paul R. Sackett, University of Minnesota. The revision of the project is being managed by Wayne J. Camara, assistant executive director for scientific affairs, APA. Diane Schneider serves as staff liaison.

1 Source: APA Public Affairs Office.

Center for Research on Evaluation, Standards, and Student Testing
Eva L. Baker, Co-director
Robert L. Linn, Co-director
Joan L. Herman, Associate Director
Ronald Dietel, CRESST Line Editor
Brenda Thomas, Layout
Katharine Fry, Editorial Assistant

UCLA
CSE/ CRESST Annex
Graduate School of Education
405 Hilgard Ave.
Los Angeles, California 90024-4108

ADDRESS CORRECTION REQUESTED EF-13