From the Directors:

A Field of Dreams
Eva L. Baker and Robert L. Linn

For the most part, validity issues of assessment, indicator, and accountability systems are studied in situ. Developed through patterns of local history, personality, regional politics, and national initiatives, many local and state assessment and accountability systems are far removed from optimal. It is often the case that when experts are invited to provide advice and critiques of plans or analyses of system operations (for example, in California, Kentucky, Colorado, New York, Kansas, and Illinois), calls for stronger validity arguments are overwhelmed by more immediate practical and political needs.

We know even as they begin that many accountability systems will struggle for a number of reasons—for instance, because poorly designed external indicators cannot survive accountability demands (Koretz, 1996; Linn, 2000; Shepard, 1993). As a result, almost no system gets it right, and too few get it close to the mark. Saying “I told you so” gives no satisfaction, and after a few post hoc cycles, researchers in assessment and accountability are labeled “naysayers.”

From technical and policy perspectives, CRESST researchers can create accountability models of what can and should be, rather than simply serving as public accountants of what has been decided. We believe that there is dual value to efforts where CRESST researchers, along with invited national and international collaborators, contribute to the design of one or more ideal assessment systems for use at state, district and classroom levels. We believe that it is important to develop the positive story, and to show how practical, technical, and economic considerations can be woven into systems that have a better chance of yielding desirable consequences. Therefore, CRESST is beginning a set of thought experiments to design an ideal system or set of ideal systems.
As a first step, we are exploring essential attributes for creating ideal accountability models. Previous CRESST work in validity principles (Linn, Baker, & Dunbar, 1991), the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999), reports by the Board on Testing and Assessment (Elmore & Rothman, 1999; Heubert & Hauser, 1999), studies of impact (Koretz & Barron, 1998; Koretz, Barron, Mitchell, & Stecher, 1996; Koretz, Mitchell, Barron, & Keith, 1996; Stecher, 1998), and policy initiatives from the various states have identified key requirements, desires, and functions of such systems. They include a number of ambitious goals, such as the measurement of progress toward desirable standards, the reporting of individual and institutional progress, the evaluation of instructional inter-
ventions, the application of feedback to improve teaching and learning, the use of results to guide policy decisions, and the use of results to ameliorate inequities.

Our first step is to begin to classify the legitimate primary and secondary purposes of assessment systems, with and without strong accountability provisions attached to them.

Third, we will identify characteristics of assessments, including their focus, provision, frequency, and reporting options. Costs will be estimated (as information rather than as an absolute constraint). We will solicit advice from an expanded set of collaborators from practice, policy and testing.

...these accountability models will encourage us to entertain new perspectives and to create compelling solutions...

At the outset, our attention will be devoted to purposes, standards and content domains, and assessments. We will have to contend with a related set of issues such as alignment, capacity, and teacher assessment and will draw on the thinking and findings from other research.

Towards Utopia? No. Thinking about the ideal assessment system in a coherent, principled way may yield greater practical insights than continually patching. Deliberately shorn of short-term political pressures and long-term traditions, these accountability models will encourage us to entertain new perspectives and to create compelling solutions for seemingly intractable assessment problems. At minimum, they will provide a positive set of examples against which local officials and citizens can compare their own systems.

As designs rather than full-blown systems, ideal system properties will provide states with a basis for comparing features of planned or existing systems and evaluating trade-offs between system goals, technical qualities, and costs. In the best case, they or significant components of the designs can lead to more appropriate solutions to problems of high standards, poor performance, and lagging capacity.
CRESST's Quality School Portfolio (QSP) project is rapidly expanding with the help of several foundations and nonprofit organizations. Consisting of two distinct computer applications, QSP provides schools a single solution for using information to make better decisions. For example:

The QSP Training Initiative involves in-depth training of school management teams, a close examination of school site plans and goals, and alignment to district objectives. The Stuart Foundation is supporting the Training Initiative in six schools in Los Angeles County while the Joyce Foundation is supporting work with six Chicago schools. In addition, educators from 40 school districts as far away as Boston have traveled to CRESST to attend a one-day intensive training workshop.

QSP funding from the North Central Regional Educational Laboratory (NCREL) and the Illinois State Board of Education is being used to provide all Illinois schools with improved evaluation tools. Additional funding from NCREL is being used to pilot test QSP in Cherry Creek, Colorado, with the goal of using technology to evaluate technology.

The National School Boards Association and the American Association of School Administrators, with funds from the U.S. Department of Education, are funding QSP to study data and reporting practices and requirements of district administrators and school board members across the country. Soliciting participation from 15 districts in nine states, this work uses cross-functional teams including school board members, superintendents, principals, and teachers. This collaboration focuses on using data for continuous improvement and public engagement.

Working with the Wisconsin Center for Educational Research (WICER), the CRESST/WICER Collaboration on Information Systems is testing the latest version of QSP in two urban schools in Milwaukee, Wisconsin, and collaborating on an R&D agenda to promote schools’ use of data to improve student learning. This collaboration is funded by the Joyce Foundation.

QSP’s initial development was funded through a cooperative agreement with the U.S. Department of Education. Please visit the new QSP Web site at qsp.cse.ucla.edu for additional information and contacts.
A comprehensive review of the quality of California education produced mixed findings, according to a new CRESST report. The authors analyzed data from many indicators including statewide Stanford 9 scores, the STAR augmentation tests, and results from the National Assessment of Educational Progress (NAEP).

Among the study's positive findings were a decreasing drop-out rate, more students taking challenging courses, and improving college entrance test scores. The negatives included low performance on tests from the National Assessment of Educational Progress and low scores for disadvantaged students on the statewide Stanford 9 achievement tests.

"Our results reflect not only the diversity in educational quality, but an assessment system that is still very much in development," said Joan L. Herman, one of the authors of the report.

"Policymakers and the media are likely to pick and choose from our results," said co-author Eva L. Baker, "but our findings indicate that student achievement in California, as in other states, is related to where you live or who you are."

The authors found that social factors of race, poverty, and English speaking ability accounted for more than 65% of the differences in performance between schools.

"The relationship is clear," said co-author Richard S. Brown. "Schools with high percentages of students with limited English speaking skills, high percentages of African American and Latino students, and higher rates of poverty related to lower average scores."

Although many other studies have supported the link between social factors and achievement, the CRESST researchers found a new concern. When not-at-risk students attended schools with high percentages of at-risk students, their performance was lower.

"What we see is disturbing," said Herman. "Student achievement should relate more to what students learn in the classroom than to the demographic make-up of their school. Unfortunately, we are not seeing the desired effect," she added.

The authors also noted the decrease in scores when non-English-speaking students were added into the state test results. The results dropped usually by at least several percentile points.

"The state policy of testing many students who are far from proficient in English," said Joan Herman, "is definitely lowering the scores."

"In California, approximately 25% of the students have limited English skills. But only 2% of the comparison group of students who took the Stanford 9 were limited English proficient students," said Herman.

Other California indicators of progress are of concern. Test scores from the National Assessment of Educational Progress, often referred to as the nation's report card, showed California students well behind students in other states in both reading and mathematics. Despite the fact that NAEP tests only fully English proficient students, California students' scores were among the lowest in the nation. On the 1996 eighth-grade mathematics assessment, California ranked 31st out of 41 states. In fourth-grade reading, California was last among 38 states taking the test. In eighth-grade writing, 20% of California students reached the proficient level compared to 25% nationally.

"California performance is low on NAEP overall," said Brown, "although at-risk students have made important improvements in math."

Brown pointed to a 7.8 percentage point increase between the 1992 and 1996 fourth-grade NAEP mathematics scores for children living in poverty as evidence of improvement.

"We have a long way to go," he added.
Positive signs in California education were significant increases in the number of students taking Advanced Placement (AP) exams. However, there are still major differences in racial percentages of students taking the tests, complicated by lack of AP course availability in many low-income areas. Although 11.2% of students are Asian statewide, they take 42.8% of the Advanced Placement Calculus examinations. Only 2.5% of African American students take the same test despite the fact that they comprise 8.8% of the state student population.

Other positive trends were reduced dropout rates and higher scores on college entrance exams. For example, dropout rates in 1988-89 were 6.6%, declining to 2.8% in the 1997-98 school year. Another positive sign was an increasing percentage of students completing the University of California or California State University entrance requirements. In 1988, 30% of students met at least one set of requirements, rising to 38.5% in 1997.

The authors emphasized that the state accountability system is undergoing massive changes intended to spur achievement and better monitor educational quality.

"The augmentation tests and the soon-to-be-developed high school graduation test," said Baker, "are designed specifically to measure the California standards." She believes that they will add important information about how well California children and schools are doing.

Baker called as well for a study of the state tests.

"Our research shows that social background is still the dominant factor on achievement, and we need to know whether California tests are capable of measuring school improvement."

A complete copy of CSE/CRESST Technical Report 519, Student Assessment and Student Achievement in the California Public School System, may be downloaded at the CRESST Web site: www.cse.ucla.edu.

New CRESST Video Program Selected as Finalist in EdPres Awards

Assessment Models, a new CRESST video program, was recently selected as a 2000 Distinguished Achievement Award finalist by the Association of Educational Publishers (EdPres).

"Being selected as a finalist," said EdPres Executive Director Charlene Gaynor, "is, in itself, a noteworthy honor."

No more than four finalists were selected by the judges in each category; in a few categories, no finalists were selected due to the rigorous judging. One category winner from among the finalists will be announced at the EdPress awards banquet on June 9, 2000, at the National Press Club in Washington, DC.

Assessment Models shows the implementation of CRESST research-based assessments in Hawaii, using the downfall of the Hawaiian monarchy as a dramatic background. Topics include the research behind the models, assessment development, administration, and scoring. Teachers describe the value of the scoring process as well as their reaction to student performance. Middle and secondary school students discuss the challenging assessments, which require them to write a comprehensive essay based on what they have learned in school and their ability to use original source materials.

Assessment Models has been accepted by the National Education Telecommunications Association for satellite broadcast to school districts and television stations around the nation. The program is 23 minutes in length and available for $15 plus shipping and tax if applicable. To order the video, please contact Kim Hurst at 310-794-9140 or kim@cse.ucla.edu. Executive producers are Eva L. Baker, Robert L. Linn, and Joan L. Herman. Produced, directed and written by Ronald Dietel.
New Yearbook Available on the CRESST Web Site

Research on Inclusion of Students With Disabilities and English Language Learners in Large-Scale Assessment

CRESST recently added a fully searchable database/Yearbook to its Web site in cooperation with the AERA Special Interest Group Research on the Inclusion of Students With Disabilities and Limited English Proficient Students in Large-Scale Assessment.

Located at www.cse.ucla.edu, the database contains comprehensive summaries of research on the inclusion of special students. Abstracts were collected from the SIG and other sources, then coded by CRESST researchers according to special criteria such as subject area, grade level, and type of accommodation.

Each abstract in the database/Yearbook is accompanied by identifying information including title, author, publication date, keywords, and details for contacting the author or organization. The Yearbook also lists:

- General information on individual state inclusion and accommodation policies;
- A form for submitting an abstract;
- A form for joining the AERA Special Interest Group.

Evaluation Programs Expand

In the past 12 months, the Center for the Study of Evaluation has continued to expand its work with over 10 active evaluations in progress. Our latest evaluations include:

- Disney Learning Partnership: Creative Learning Communities, a grant program that provides funding to public elementary schools for the design and use of creative teaching strategies to improve student performance;
- The LA's BEST After School Enrichment Program, a widely recognized program to stem the problems of poor, inner city students;
- The State-Wide Evaluation of Outreach Efforts: A Value-Added Analysis. CSE researchers are evaluating the University of California's efforts to improve high school students' competitive eligibility to attend the University of California.

For more information about CSE evaluation projects, contact CRESST Associate Director Joan L. Herman at 310-206-3701 or herman@cse.ucla.edu.

CRESST Line
Recent CRESST Reports

The following reports are available from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). To order any CRESST publication or for a complete catalog, contact Kim Hurst at 310-794-9140; email, kim@cse.ucla.edu; or CRESST/UCLA, GSE&IS Building, Mailbox 95122, Los Angeles, CA 90095-1522. Most CRESST publications are also available at no cost on the CRESST Web site, www.cse.ucla.edu.

Practical Deliberation in Local School Districts: A South Carolina Experiment
Lorraine M. McDonnell and M. Stephen Weatherford
CSE Technical Report 520, 2000 .......... $5.50

Student Assessment and Student Achievement in the California Public School System
Joan L. Herman, Richard S. Brown, and Eva L. Baker

Bayes Nets in Educational Assessment: Where Do the Numbers Come From?
Robert J. Mislevy, Russell G. Almond, Duand Li Yan, and Linda S. Steinberg

The Role of Classroom Assessment in Teaching and Learning
Lorrie A. Shepard
CSE Technical Report 517, 2000 .......... $7.00

Assessing Active Knowledge
Robert Glaser and Carl P. Baxter

The Dependability and Interchangeability of Assessment Methods in Science
Noreen M. Webb, Jonah Schlackman, and Brenda Sugrue
CSE Technical Report 515, 2000 .......... $4.00

“No Excuses”: School Reform Efforts in Exemplary Schools of Kentucky
Shelby Wolf, Hilda Borko, Monette McIver, and Rebekah Elliott

Developing Indicators of Classroom Practice to Monitor and Support School Reform
Pamela R. Aschbacher

Professional Development: A Key to Kentucky’s Reform Effort
Hilda Borko, Rebekah Elliott, and Kay Uchiyama

Identifying Differential Item Functioning on the NELS:88 History Achievement Test
Vi-Nhuan Le
CSE Technical Report 511, 1999 .......... $2.50

Accuracy of Year-1, Year-2 Comparisons Using Individual Percentile Rank Scores: Classical Test Theory Calculations
David Rogosa

Accuracy of Individual Scores Expressed in Percentile Ranks: Classical Test Theory Calculations
David Rogosa
CSE Technical Report 509, 1999 .......... $5.50

Accommodations for Students With Disabilities: A Teacher’s Guide
Ann M. Mastergeorge and Judy N. Miyoshi
CSE Technical Report 508, 1999 .......... $4.00

Knowledge Mapping in the Classroom: A Tool for Examining the Development of Students’ Conceptual Understandings
Ellen Osmundson, Gregory K. W. K. Chung, Howard E. Hert, and Darina C. D. Klein

State Standards-Setting and Public Deliberation: The Case of California
Lorraine M. McDonnell and M. Stephen Weatherford
CSE Technical Report 506, 1999 .......... $5.00

Spring 2000
Congratulations to CRESST Partner
Lorrie Shepard

C O N G R A T U L A T I O N S  t o Lorrie Shepard, who recently completed her tenure as president of the American Educational Research Association (AERA). During the recent AERA annual meeting in New Orleans, Creating Knowledge in the 21st Century: Insights From Multiple Perspectives, Dr. Shepard delivered the presidential address to a packed audience.

With national attention increasingly focused on raising school accountability through student testing, Shepard called for a fundamental change in how student learning is assessed.

“Our goal should be to find ways to fend off the negative effects of externally imposed tests and to develop instead classroom assessment practices that can be trusted to help students take the next steps in learning,” she said. “Tests are only one way to assess student learning,” Shepard added.

She also called for more opportunities for teachers to learn about assessment in order to improve teaching and learning.

CRESST extends its condolences to the family and friends of Frederic Mather Lord, a principal developer of the statistical machinery underlying modern mental testing, who died on February 5, 2000, in Naples, Florida. He was 87. Dr. Lord was recognized as a preeminent authority in the statistical theory of mental tests.

In March of 1949 Dr. Lord joined the newly formed Educational Testing Service as Head of Statistical Analysis. In 1950 he moved to the Research Division where he remained until his formal retirement in 1982. In addition to his wife, he is survived by his sons, John (of Palmetto, Florida) and Eric (of Langhorne, Pennsylvania), his three stepchildren, Conrad M. Bessemer and Robert H. Bessemer (of Pittsburgh) and Diane Carnegie Bessemer (of Seattle), and grandchildren. We will miss his ongoing contributions to the measurement field.