**NCLB** is on everyone’s mind. Recently drafted regulations emphasize the importance of publicly available validity data for each of the purposes for which state accountability assessments are used. Such provisions are consistent with core tenets of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) and the CRESST/CPRE Standards for Educational Accountability Systems (Baker & Linn, in press) and might well kindle hope that high-quality assessments will be one of the central results of the No Child Left Behind (NCLB) enactment.

In order to meet these provisions, a state will need to take a number of steps.

1. Make explicit the purposes that the assessment system and individual assessments are intended to serve.
2. Decide on a strategy to meet the testing requirements at various grade levels.
3. Determine the degree to which validity evidence is available or could be accumulated for multiple purposes and “the widest possible range of students.”
4. Determine a standard of adequacy for technical quality.
5. Make plans to acquire needed technical quality information during piloting, field trials, and implementation.

These are substantial requirements. Let’s consider just one: the purposes of the assessment. State assessments have multiple intended uses, but accountability is regarded as preeminent. The use of assessment data to classify schools in categories (e.g., needs improvement, satisfactory, eligible for a reward) is a bookkeeping function, where a school’s status and progress are tabulated annually against particular targets. Lots of different measures could be used to classify schools. The important point here is if schools are expected to do something about the data, then the details of how the assessments are designed and technically vetted matter.
Test-based accountability is founded on the premise that once the information is made available from the assessment system, reasoned decisions will yield subsequent improvement in performance. Putting aside for the moment important issues of infrastructure, instructional resources, and teacher quality, for this outcome to occur there are minimum technical requirements for the assessments used in the system. A primary one is that the test(s) serving as an information source in the accountability system will detect, if present, growth attributable to instruction and school services in the target standards. Meeting this provision requires a multifaceted approach, as there are a number of ways that results can give rise to inaccurate inferences:

- If performance is flat (growth targets are not met), it may be because the test underrepresents standards and instruction occurring in the school, a case of construct underrepresentation. Good things are happening, the school is addressing important standards, and students are learning, but the test doesn’t pick it up.
- If performance is flat, it may be because the test has attributes that are irrelevant to domain performance and instruction (construct irrelevance). The test is measuring background characteristics more than the impact of educational services.
- If performance is flat, it may be because inadequate instruction has occurred, and the assessment system does not include indicators of instruction and good alignment practice.
- If performance is flat, it may be because good instruction has occurred, but the test is insensitive to measuring change.
- If targets are met, it may be because scores are inflated (false positive) through test preparation and teaching content measured only by the unique combination of format and content on the selected test.

We need to get these inferences straight in order to be confident that we are measuring the standards intended and appropriately providing assistance to schools that need to improve, and rewarding schools that deserve it.

Validity information regarding the above points is not generally available from commercial developers. However, states can take action by requesting testing contractors or independent evaluators to obtain data or conduct studies that investigate these features (e.g., construct-representation and construct-irrelevance of the test, instructional sensitivity, and score inflation). These studies should use representative samples and should be conducted separately for measures focused on content, for those with heavy language components, and for subgroups. Such studies can be incorporated into the procurement process and need to occur as early as possible. Some should be incorporated into the test development or selection process, others can occur during field testing, and others should extend throughout the operational period. For example, focused studies of narrowed preparation can be conducted experimentally, but may await the larger field tests of new assessments. Teacher surveys may be used to ask about content covered by instruction, although such work has
yielded uneven results. Such studies will need to include good measures of test alignment, as well as added sets of items that are intended to measure various degrees of generalization (format, content knowledge, application contexts), to determine whether standards have been achieved.

One practical tension is very clear: Approaches that solve the accountability problem (general, broad sample, many items, indexing) are very likely to exacerbate the instructional sensitivity problem. Broad-based measures are far less likely to reflect instruction. In any case, these accountability and instructional improvement purposes are the minimum required by the legislation, and these purposes should guide the approaches taken to study the validity of systems.

It cannot be overemphasized: If tests used in the system are not demonstrated to measure growth inspired by instruction, they are of no use to the accountability system and will mislead teachers, educators, policymakers, and the architects of NCLB.

Validity studies should assess whether, once having test results available, schools and teachers are advantaged by the information that they received, either for targeting particular efforts, redesigning instruction, or developing new patterns of services. To investigate the consequences of the availability of accountability information, both the analysis of test properties and the investigation of the likely impact of various improvement strategies or remedies are needed.

Additional purposes, such as the placement and promotion of individual students, imply careful analyses of other consequences of test use and a different reason for investigating the boundaries of achievement levels. Communication functions of measures (to determine the degree to which teachers make appropriate inferences from performance) can also be addressed, but may be incorporated in the larger question of alignment.


CRESST 2002 Conference

This year the CRESST Conference is being held at the newly renovated Radisson Hotel, Los Angeles Airport. Conference dates are Tuesday, September 10–Wednesday, September 11, 2002. Complete details and registration information are on the CRESST Web site: www.cse.ucla.edu

Presenters at this year’s conference are:

Jamal Abedi, CRESST/UCLA
Jane Armstrong, Education Commission of the States
Lyle Bachman, UCLA
Alison Bailey, CRESST/UCLA
Eva L. Baker, CRESST/UCLA
Hilda Borko, CRESST/University of Colorado at Boulder
Michael Brown, UC Santa Barbara
Wayne Camara, The College Board
Joseph Conaty, U.S. Department of Education
Robert Cooper, CRESST/UCLA
Pasquale DeVito, National Research Council
Stephen Dunbar, University of Iowa
Mark Feller, California Department of Education
Robert Glaser, CRESST/University of Pittsburgh
Brian Gong, The National Center for the Improvement of Educational Assessment
Edward Haertel, CRESST/Stanford University
Joan L. Herman, CRESST/UCLA
Harold Himmelfarb, OERI, U.S. Department of Education

(see CRESST Conference, page 5)
A new report from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA raises concerns about the way in which states may choose to implement growth targets required by the No Child Left Behind Act of 2001 (NCLB).

The Act requires that all students must reach at least the proficient level on their state exams by the year 2014. Each year, schools must increase their percentage of students at the proficient or above level by an amount sufficient to move from their 2002 baseline level to 100% over the next 12 years. Most states now report scores at four achievement levels, such as below basic, basic, proficient, and advanced.

“The policymaker expectations almost certainly exceed the ability of schools to make this sort of progress,” said one of the study’s authors, Robert Linn, a CRESST co-director and current president of the American Educational Research Association. The “proficient” level in most states is set very high.

The CRESST researchers examined past state achievement results in several states, as well as the National Assessment of Educational Progress (NAEP) results for those states. In Colorado, for example, using a standard of a 1% increase in the percent proficient or higher—an amount much less than would be required by NCLB—they found that only one school in 20 would have met the target increase of one percentage point in each of three years. This is despite the fact that the average yearly increase was more than 1.5% per year.

“One sticking point,” said co-author Eva Baker, “is that the new law requires steady increases from year to year between now and 2014.” A school can increase sufficiently to meet the target in one year, but if it slips below the target in subsequent years it can fall into the “needs improvement” category. Baker is a CRESST co-director and chairs the National Research Council’s Board on Testing and Assessment.

Linn added that scores at the school level are very volatile from one year to the next, making substantial and steady annual progress difficult. Measurement and sampling error contribute to the stability problem. The researchers predicted that a substantial number of schools will be incorrectly identified as needing improvement as a result of the volatility of school-level scores.

The authors also found that the goal of having all students reach the proficient or above level by 2014 was highly dependent on the difficulty of the state test or where the achievement levels were set. Their comparisons showed that in the year 2000, only states such as Texas, which has basic skills tests and relatively modest standards, had a large enough proportion of their students meeting standards for the 100% proficient or above goal to be within the realm of what might be possible in 12 years. However, Texas is revising its standards as well.

In states with more demanding tests and more stringent standards it is not unusual to have less than 2 students in 5 scoring at the proficient level or higher. For example, in 2000,
only 30% of Maryland students scored proficient or above on their state test, the Maryland School Performance Assessment Program.

To verify the Texas and Maryland results, CRESST researchers compared them to the National Assessment of Educational Progress. In Texas, only 25% of students achieved at the NAEP proficient level or above, whereas in Maryland, the figure was 28%. Under the new law, NAEP will be used to validate performance increases for all states receiving federal Title I funds.

The analyses confirm major differences between state tests. The results show that many states are at very different starting places and will have very different distances to go in order to meet the law's requirements (Figure 1).

Based on the state NAEP results, the researchers concluded that moving 70% or more of all students to a proficient or above level is a worthy, but not realistic, goal. Further, in analyses of state NAEP reading scores from 1992 to 1998, only 3 of 33 states met a standard of a 1% annual increase, which is far below the 5% or 6% per year increase that would be required to reach 100% in 12 years.

The problem is exacerbated by the requirement that all subgroups of students reach the proficient or above level and increase at the required rate, including students with disabilities and English language learners. If the overall school percent proficient increases enough to meet the adequate yearly progress target, but students with disabilities or English language learners fail short of the target, the school will have failed to meet the target.

The NCLB goals are laudable, concluded the researchers, but even the NAEP basic achievement level would be a challenging goal for most schools to reach in 12 years. The authors recommended a switch from the use of achievement levels, which have had a history of challenges, to more traditional methods for reporting progress.

The complete report can be found on the CRESST Web site at http://www.cse.ucla.edu.


CRESST Conference (from page 3)

Tyrone Howard, CRESST/UCLA
Thomas Kane, CRESST/UCLA
Daniel Koretz, CRESST/Harvard University
Robert L. Linn, CRESST/University of Colorado at Boulder
Scott Marion, Wyoming Department of Education
Wayne Martin, Council of Chief State School Officers
Heinrich Mintrop, CRESST/UCLA
Robert Mislevy, CRESST/University of Maryland
William Padia, California Department of Education
Mari Pearlman, ETS (Invited)
James Pellegrino, University of Illinois, Chicago
James Popham, UCLA
Lauren Resnick, CRESST/University of Pittsburgh
Bella Rosenberg, American Federation of Teachers (Invited)
William Sanders, SASinc and University of North Carolina
Cynthia Schmeiser, ACT, Inc.
Susan Scalfani, U.S. Department of Education
Lorrie Shepard, CRESST/University of Colorado at Boulder
Brian Stecher, CRESST/RAND
Mick Walker, Qualifications and Curriculum Authority, United Kingdom
CRESST is very pleased to announce the following new partners. Each researcher has unique skills that will further our mission to conduct research leading to the effective assessment of educational quality.

**Abeer Alwan**  
Professor, UCLA Henry Samueli School of Engineering and Applied Science  
Dr. Alwan’s research includes modeling human speech production and perception mechanisms and applying these models to speech-processing applications such as automatic recognition, compression, and synthesis. She is a recipient of the NSF Research Initiation Award (1993), the NIH FIRST Career Development Award (1994), the UCLA-TRW Excellence in Teaching Award (1994), the NSF Career Development Award (1995), and the Okawa Foundation Award in Telecommunications (1997). She received her PhD in electrical engineering from the Massachusetts Institute of Technology in 1992.

**Robert Cooper**  
Assistant Professor, UCLA Graduate School of Education & Information Studies  
Dr. Cooper conducts research on the implementation and scale-up of school reform models. His research focuses on the politics and policies of school reform, particularly as they relate to issues of race and equity for at-risk students. Specializing in the use of a mixed-methods approach, he has published and presented numerous papers on the varying aspects of school reform and school change, including recent articles in *Urban Education*, the *Journal of Negro Education*, *Education and Urban Society* and the *Journal of Education for Students Placed at Risk*. He holds a PhD from UCLA.

**Margaret Heritage**  
Principal, Corinne A. Seeds University Elementary School, and Senior Researcher, CRESST/UCLA  
Margaret Heritage is transitioning to become a full-time CRESST senior researcher from her current position as principal of the Corinne A. Seeds University Elementary School at UCLA, where she led the school to numerous program improvements. In the United Kingdom, she taught elementary school for several years and was a principal before taking a position as a county education inspector for elementary education serving 250 schools. In addition to her work in schools, she has participated in a wide variety of research initiatives and published on topics including literacy, children’s personal and social development, curriculum development, and school management.

**Tyrone Howard**  
Assistant Professor, UCLA Graduate School of Education & Information Studies  
Tyrone Howard completed his doctoral work at the University of Washington where his specialty area included multicultural education and social studies in education. His research focuses on issues of equity and access in urban schools and the education of African American students. He is an active participant in UCLA’s Teacher Education Program and was previously an assistant professor at Ohio State’s College of Education.

**Thomas Kane**  
Professor, UCLA School of Public Policy and Social Research  
Thomas J. Kane is a Professor of Policy Studies and Economics at UCLA. He has studied a number of issues related to higher education: estimating the labor market payoff of a community college education, observing the
impact of tuition and financial aid policy on college enrollment rates, and analyzing the impact of affirmative action in college admissions. He is currently working on an evaluation of the Cal Grant program on college enrollment rates and on the design of K-12 school accountability systems.

Prior to coming to UCLA, Professor Kane was an Associate Professor at the Kennedy School of Government at Harvard University. He served as the senior economist for labor, education and welfare policy issues for President Clinton’s Council of Economic Advisers. Dr. Kane received his PhD in public policy from Harvard University.

Heinrich Mintrop
Assistant Professor, UCLA Graduate School of Education & Information Studies

Heinrich Mintrop’s research has focused on issues of educational policy, school accountability, school improvement, democratization, and cross-national studies. He received a Carnegie Corporation scholarship to compare school accountability systems in the United States and Germany. Prior to his university career, he was a teacher in the United States and Germany. He received his PhD in education from Stanford University.

Victoria Vesna
Chair, UCLA School of Arts and Architecture

Victoria Vesna came to UCLA from the University of California, Santa Barbara, where she was a professor of art. Her work has ranged from performance and video installations to experimental research that connects networked environments to physical public spaces. She explores how physical and ephemeral spaces affect collective behavior. Vesna is the North American editor of Artificial Intelligence & Society and is working on a special issue, “Database Aesthetics: Issues of Organization and Category in Art.”

Recent CRESST Reports

The following technical reports can be downloaded from the CRESST Web site, www.cse.ucla.edu, or ordered by e-mail from Kim Hurst: hurst@cse.ucla.edu.

Benchmarking and Alignment of Standards and Testing
Robert Rothman, Jean B. Slattery, Jennifer L. Vranek, and Lauren B. Resnick
CSE Technical Report 566, 2002 ............ $4.00

Combining Surveys and Case Studies to Examine Standards-Based Educational Reform
Brian Stecher and Hilda Borko

Teacher Effects as a Measure of Teacher Effectiveness: Construct Validity Considerations in the TVAAS (Tennessee Value-Added Assessment System)
Haggai Kupermintz
CSE Technical Report 563, 2002 ............ $3.00

Looking Into Students’ Science Notebooks: What Do Teachers Do With Them?
Maria Araceli Ruiz-Primo, Min Li, and Richard J. Shavelson
CSE Technical Report 562, 2002 ............ $3.50

Stability of School Building Accountability Scores and Gains
Robert L. Linn and Carolyn Haug
CSE Technical Report 561, 2002 ............ $2.00

Examining Relationships Between Where Students Start and How Rapidly They Progress: Implications for Conducting Analyses That Help Illuminate the Distribution of Achievement Within Schools
Michael Seltzer, Kilchan Choi, and Yeow Meng Thum
CSE Technical Report 560, 2002 ............ $4.25

Latent Variable Modeling in the Hierarchical Modeling Framework: Exploring Initial Status Treatment Interactions in Longitudinal Studies
Michael Seltzer, Kilchan Choi, and Yeow Meng Thum
CSE Technical Report 559, 2002 ............ $3.50
We are saddened to learn of the death of Robert Gagné on April 28, 2002. Dr. Gagné contributed to the Center for the Study of Evaluation’s early years and was a leader, along with CRESST partner Robert Glaser, in the development of the concept of systematic instruction.

“Robert was a key influence on education,” said CRESST Co-director Eva Baker. “He will be remembered for his contributions to educational evaluation, especially his research on task analysis and his development of a taxonomy of learning, which is still in wide use today.”

Gagné was also known for his seminal work, The Conditions of Learning, and his tremendous influence on science education.

Dr. Gagné was a fellow of the American Association for the Advancement of Science and held faculty positions at Connecticut College for Women (1940), Princeton (1958–1962), UC Berkeley (1966–1969), and at Florida State University from 1969 until his retirement in 1985. He was the director of research at the American Institutes for Research from 1962–1966.

Robert Gagné: In Memoriam

CRESST and its parent center, the Center for the Study of Evaluation, recently competed for and won several major awards, totaling nearly $15 million.

With $4.7 million from the Interagency Educational Research Initiative (IERI), CRESST will design the Assessment Design and Delivery System (ADDS) to support teacher assessment in science. Using research-based assessment models, this system will assist teachers in understanding and assessing student progress in science course topics. The grant is spread over 4 years.

The Office of Naval Research is funding CRESST approximately $7.4 million across 3 years for the development of an integrated set of distance learning guidelines and assessment models. While primarily supporting Navy training curricula, the research will create multiple products useful to other military, governmental, and corporate organizations that are also making large investments in distance learning technology.

Finally, the U.S. Department of Education is funding CRESST over $2.7 million during 3 years to create an online version of the Quality School Portfolio. QSP is already used in 49 states to analyze data and school quality. The QSP Web site will allow teachers to use the information for instructional improvement and parent conferences.