



FROM THE PRESIDENT

By Suzanne Lane, University of Pittsburgh

I am very pleased to announce that Michael Kolen has accepted the appointment as the editor of the *Journal of Educational Measurement* as of January, 2005. Mike will be working closely with Barbara Dodd during the 2004 year, and as of January 2004, all submissions to *JEM* should be sent to Mike. Further, the slate for NCME Board positions is very impressive: James Impara and Carole Perlman for Vice President, James Carlson and Linda Cook for Board Member at Large, and Terry Ackerman and Barbara Dodd for Board Member from a University. Ballots will be sent this month so remember to cast your votes!

The NCME Board and Committees are working on a number of projects that are in response to the continued attention to educational testing. For example, as reported in the last issue of the *Newsletter*, the NCME Outreach Committee is building relationships with other educational organizations, including the National Council of State Legislators (NCSL), American Association of School Administrators (AASA), and National Association of Secondary School Principals (NASSP). There will be a session at the 2004 annual meeting that will continue a dialogue among representatives from these three organizations and NCME members. Please look for it in our Annual Meeting Program.

With the increased attention to testing, we need many more professionals in our field. The Board will be forming an ad hoc committee to propose federal support for training in educational assessment and measurement. Further, the NCME Recruitment Committee is developing a brochure aimed at recruiting students into educational measurement and statistics graduate programs. They are also developing short biographies of professionals in our field to highlight the important work that we all do. These biographies will be on our website within the coming months.

NCME membership has been declining!! In 1997, we had 2,300 members, and 6 years later our membership has decreased to 1,751. This is an extremely puzzling phenomenon given the increasing role of testing in our nation's educational system. A major goal of the Board is to increase membership so as to ensure that we are having a dialogue with the many thousands of professionals that are involved in developing, using, and/or interpreting tests. We need your help, however. For example, inform your colleagues about NCME, invite them to contact our website (www.ncme.org) where they can obtain additional information and a membership form, and encourage them to become members. Don't assume that your colleagues are already members. I just glanced at the names of our 1,751 members and was amazed that many of our colleagues across the nation are not current members of NCME. Also, invite your students to become members. You can download membership forms from our website and distribute them in your classes. The fee is only \$60 for professionals and \$30 for students and the fee includes our two excellent journals, *Journal of Educational Measurement* and *Educational Measurement: Issues and Practice*. The NCME Membership Committee has been active in developing recruitment strategies for new members. Last year, they were successful in recruiting members from CCSSO and this year they are inviting board members from the National Council of State Legislators (NCSL), American Association of School Administrators (AASA), and National Association of Secondary School Principals (NASSP) to join NCME.

Carol Parke and Sue Brookhart have an impressive set of invited sessions for our Annual Meeting in April, including *Hierarchical Modeling of Social and Cognitive Assessment Data*, and *Meshing Measurement with Curriculum and Instruction: Three Venues, Three Strategies*. Allan Cohen has also identified a number of excellent training sessions for the meeting. We are fortunate to have such competent colleagues working on our annual meeting!

NCME's financial standing is an area of concern. The Board has identified a number of revenue generating ideas. After a careful review of annual meeting registration fees charged by other organizations, we decided to raise the registration

continued on page 2

continued from page 1

fees for our Annual Meeting as of 2004. Pre-registration fees were raised to \$60 for NCME members, \$30 for students, \$90 for nonmembers and on-site registration fees were raised to \$75 for NCME members, \$35 for students, and \$100 for nonmembers. These fees are still considerably lower than other organizations' annual meeting fees. Secondly, as of the June 2004 issue, the NCME newsletter will be published on-line only. In the April Board meeting we will continue to discuss other revenue generating ideas such as posting paid advertisements for positions on our website and offering an online program for our future meetings. The Central Office is identifying ways to enhance the efficiency of NCME administrative activities and to promote the products of NCME. The outcome of one initiative is that NCME will change to an online membership service, allowing members to update address information and renew memberships online. We hope to have this in place sometime during the 2004 year. The Central Office is also responsible for developing a CD-Rom for *JEM* that contains the first issue through the final issue of 2000. It has the capability to search for articles by title, author or keyword, and articles can be downloaded as they appeared in the journal. Please see the order form in this newsletter.

Lastly, NCME's contract with AERA management services expires in June 2005. The Board is gathering information about continuing with the Central Office as well as alternative models for handling NCME's administrative responsibilities, annual meeting, and publication of *EM:IP* and *JEM*.

I welcome your comments and suggestions on any of the above topics. I can be reached by e-mail at sl@pitt.edu.

CASMA-ACT CONFERENCE ON CURRENT CHALLENGES IN EDUCATIONAL TESTING

By Stephanie JL Gertz, McDougal Littell

ACT and The University of Iowa's Center for Advanced Studies in Measurement and Assessment (CASMA) hosted a one-day conference in Iowa City on the subject of "Current Challenges in Educational Testing" on November 8, 2003. Approximately 175 people attended the meeting. It was a unique opportunity to listen to some of the leaders in our field and have the chance for interaction in a small setting.

The agenda was full, starting with a welcome by Robert Brennan (CASMA Director) and welcome from Cynthia Schmeiser (Senior Vice President ACT) and Sandra

Damico (Dean College of Education, University of Iowa). Speakers included: Robert Linn, University of Colorado and CRESST; Wayne Camara, College Board; Cynthia Schmeiser, ACT; Ronald Berk, Johns Hopkins University; Randy Bennett, Educational Testing Service; Richard Stiggins, Assessment Training Institute; Steve Kromer, NCS; John Laramy, Riverside; and John Oswald, ETS.

CASMA is a new center at the College of Education at the University of Iowa. Its primary purpose is to pursue research-based initiatives that lead to advancements in the methodology and practice of educational measurement and assessment. This was the first conference; subsequent conferences may be held on a biennial basis. The Center was started in September, 2002. Robert L. Brennan, the Lindquist Chair in Measurement and Testing at the University of Iowa, is the Director. For more information, visit the website (www.uiowa.edu/~casma/mission.htm).

NEWSLETTER ADVISORY BOARD

ROBERT ANKENMANN, University of Iowa
JUDY ARTER, Assessment Training Institute
SCOTT BISHOP, Riverside Publishing
MICHELINE CHALHOUB-DEVILLE, University of Iowa
KATIE FISK, Connecticut State Department of Education
JOAN HERMAN, CRESST/UCLA
SHARON LEWIS, Council of the Great City Schools
DUNCAN MACQUARRIE, Harcourt Educational Measurement
WENDY MCCOLSKEY, SERVE
HILLARY MICHAELS, CTB/McGraw-Hill
CAROL S. PARKE, Duquesne University
S.E. PHILLIPS, Consultant
BARBARA PLAKE, Buros Center for Testing, University of Nebraska-Lincoln
DOUGLAS RINDONE, Harcourt Educational Measurement

SUSAN M. BROOKHART, Editor, Duquesne University

Send articles or information for this newsletter to:

Susan M. Brookhart Phone: (406) 442-8257
2502 Gold Rush Avenue Fax: (406) 442-8257
Helena, MT 59601 e-mail: susanbrookhart@bresnan.net

The *NCME Newsletter* is published quarterly. The *Newsletter* is not copyrighted; readers are invited to copy any articles that have not been previously copyrighted. Credit should be given in accordance with accepted publishing standards.

NEWSLETTER TO GO ALL-ELECTRONIC

In October, the NCME Board voted to make the Newsletter available in electronic format only, starting with Vol. 12, No. 2 (June, 2004). Issues of the Newsletter will be posted on the NCME website (www.ncme.org). Notice of new issues will be sent via the NCME Listserv. Information about subscribing to the listserv is also on the website.

THE SOUTH CAROLINA READINESS ASSESSMENT FOR KINDERGARTEN AND GRADE ONE

*By Huynh Huynh, University of South Carolina, and
Theresa Siskind and Jim Casteel, South Carolina
Department of Education*

South Carolina sought a developmentally appropriate assessment system that could be used with kindergarten and first grade students to help determine their readiness for school. Children in kindergarten and the lower grades grow and develop a great deal in a short period of time. Growth often varies among children and does not always occur in a linear fashion. Recent research has shown the value of measuring a child's growth and learning in a more continuous way. Models of continuous assessment provide documentation for a child's growth in a number of areas over time and can be more useful than traditional assessments to determine a child's readiness for school. The Cognitive Skills Assessment Battery (CSAB) was used in the state of South Carolina starting in 1978. Administered in the first 15 days of first grade, it provided a single measure of readiness for school. The CSAB was used for the last time in fall 2001 and was replaced by the new South Carolina Readiness Assessment (SCRA) in 2001-02.

The path to this transition began in 1993, with the passage of the Early Childhood Development and Assistance Act (Act 135). Three continuous assessment systems were piloted during this year: the Primary Level Assessment System (PLAS), the Work Sampling System (WSS), and some independent assessment systems developed by individual schools. The WSS was well received, and the system was pilot-tested during the following year with a larger number of schools. The Education Accountability Act of 1998 reinforced this direction when it explicitly required readiness assessments for grades 1 and 2 that were "not to be used as an accountability measure."

In 1999, the South Carolina State Department of Education (SDE) extended a call for proposals for a statewide continuous assessment system that would provide a better understanding of student readiness for school. Harcourt Educational Measurement submitted a proposal to use a version of the WSS modified specifically for the state of South Carolina. The system was aligned with the state curriculum standards in mathematics and English language arts for kindergarten and first grade. The first statewide field test of this system took place during the 2000-01 academic year. The system involved two checklists, one for kindergarten students and one for those in first grade, and development guidelines to help teachers have similar understandings of the checklists. Minor revisions were made to the system from results of this first pilot test and the SCRA was field tested statewide during the 2001-02 year. The Work Sampling System (WSS) is a curriculum-embedded continuous assessment process. Teachers observe their students during every day classroom activities to gain a more complete picture of the students' development. The WSS was originally developed by a team at the University of Michigan headed by Samuel J. Meisels and is designed for use with students in kindergarten through fifth grade.

Many groups are involved with the project., and the Office of Assessment of the SDE serves as leader and coordinator. Support for educators is provided by the Office of Early Childhood. The service provider is Harcourt Educational Measurement. Rebus, Inc., the company that developed the WSS, is now owned by Pearson Educational Measurement. Harcourt administrators work with the SDE to produce the checklists and developmental guidelines. They also manage the scoring and reporting processes. An SCRA Advisory Committee was also established in November 2001. This group consisted of teachers and administrators from schools around the state, SDE administrators, and personnel from both Rebus, Inc., and Harcourt Educational Measurement. The Advisory Committee has met regularly and has recommended a number of substantial changes to the SCRA. The SDE administrators also work with the SCRA advisory committee, a group of teachers and administrators from schools and districts around the state, to gain valuable feedback from those directly involved with the assessment. The Education Oversight Committee has the responsibility of reviewing the fully-developed assessment program to ensure it meets the goals of the state and the needs of South Carolina students.

continued on page 4

continued from page 3

The general structure of SCRA consists of two complementary elements: developmental guidelines and checklists, and summary reports. The current checklist consists of three domains (Personal and Social Development, English/language arts, and Mathematics), 14 functional components, and approximately 40 indicators. The indicators vary according to grade level. A minimum of two assessments (winter and end-of-year) of student performance on curriculum standards are made during an academic year. Only the final yearly assessment is used for reporting performance. Initially all teachers were trained to use the checklists during summer workshops. A train-the-trainer model is now employed with additional professional development activities provided by the Office of Early Childhood.

Using scannable checklists or an online system developed by the State, teachers make two, three, or four assessments (at district direction) of student performance on curriculum standards during an academic year (fall, winter, spring, and year end). During each assessment period, students are rated on their demonstration (consistently, sometimes, rarely or never) for each of the indicators. The scannable checklists and computer file from the online system are sent to the Harcourt San Antonio Scoring Center for scoring during the three weeks before the end of the school year. Upon completion of scanning the checklists and scoring the functional components, individual student reports, class rosters, and school/district/state summaries are produced and shipped to districts. The district test coordinator is responsible for ensuring that schools receive the Student Home Report and Class Roster at the beginning of the school year.

An online version of the SCRA was piloted in 2002-03. Approximately one-third of the student checklists were completed online. It is anticipated that three-fourths of the checklists will be completed online in 2003-04. The online system is preloaded with teacher and student information from the SDE database. Teachers logon to the secure website from any location with Internet access. The online screens mimic the appearance of the scannable checklists with click-on buttons in place of pencil bubble-in. The data are continuously uploaded to the SDE preventing the need for districts to submit checklists at the end of the year.

Technical documentation for the SCRA was written by Huynh Huynh, Susan Prior, and Dorinda Gallant. It will soon be available in the website of the South Carolina Department of Education (MySCschools.com).

NEW CRESST POLICY BRIEF ON ENGLISH LANGUAGE LEARNERS

By Ron Dietel, CRESST/UCLA

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) recently completed a policy brief focusing on the issues surrounding English Language Learners (ELLs) in meeting the requirements of the No Child Left Behind Act. *Challenges in the No Child Left Behind Act for English Language Learners*, by Jamal Abedi and Ron Dietel, may be found on the CRESST web site (CRESST.org) under Policy Briefs. The policy brief identifies ten special challenges for the ELL subgroup (below) and suggests improvements for the assessment of ELL students.

1. Historically low ELL performance and very slow improvement – CRESST research, supported by NAEP and state test results, shows that English Language Learners consistently perform lower than other students and frequently lower than any other subgroup. Rapid progress by students overall combined with policies that test ELL students who have lived in the United States for very short periods of time contribute to a growing ELL achievement gap in many states and school districts.

2. Measurement accuracy – CRESST studies have repeatedly shown that English Language Learners perform substantially lower on language arts tests compared to mathematics and science tests. Studies measuring accommodations effects have improved ELL performance by approximately 10-20% on many tests. Modifying, often simplifying, the language of the test items (see, for example, Abedi & Lord, 2001), has consistently shown ELL performance improvement without reducing the rigor of the test. These findings suggest that low ELL language ability decreases ELL performance on most tests, thus influencing the test as an accurate measure of ELL content knowledge. The test becomes a measure of two skills for the ELL student, language *and* subject.

3. Instability of the ELL student subgroup – Researchers have long postulated that a central cause of flat ELL test scores is the regular removal of high achieving students from the ELL subgroup. Once ELL students become language proficient, they are redesignated fully English proficient (RFEL) and removed from the ELL subgroup. In a study of approximately 14,000 students, CRESST found that redesignation and removal of students from the main ELL group coincided with a modest but significant performance drop by ELL students in reading. While

continued on page 5

continued from page 4

this is not a claim of causation between ELL redesignation and dropping scores, it reinforces the redesignation dilemma specific to the ELL subgroup. States, districts, and schools with a growing ELL population face an added challenge. A constant increase of low achieving ELL students, even if all other factors remain constant, will make it more difficult to achieve adequate yearly progress for both the ELL subgroup and the overall student population.

4. Factors outside of a school's control – CRESST research supports hundreds of other studies which show that non-school factors, usually parent educational level or socioeconomic status, outweigh school factors on student achievement. A study of more than 30,000 students found these influences operate even within the ELL population itself. In this study, the gap between ELL students whose parents had postgraduate education compared to students whose parents had not graduated from high school was approximately 15 percentile points. Non-school factors are strong and persistent.

5. ELL identification – NCLB is helping to create a more common definition of ELL students, but accurate ELL identification remains a challenge. CRESST research has found that many school ELL identification methods produce inconsistent results. These studies have found a weak relationship between ELL classification codes and language proficiency test scores or achievement test scores.

6. Subgroup size – NCLB does not set a minimum number of students necessary to require subgroup reporting, but requires large enough numbers to provide statistically reliable data. A low number, say 25, will increase the number of schools reporting ELL subgroups but have the negative effect of more variable performance (Linn, Baker, & Herman, 2002). Setting a higher minimum number of students, say 100 (Hill & DePascale, 2003), will produce less performance variation and thus a more dependable growth rate, but will substantially reduce the number of schools that must report English Language Learners as a subgroup. Allowing states to have different minimum group sizes will create comparability and fairness issues between state accountability systems.

7. ELL subgroup diversity – Another challenge to states is the diversity of student achievement within ELL subgroups. ELL performance varies along racial lines, for example, similar to performance differences in the general student population. In one study, we found substantial differences in performance between ELL students with a Chinese-speaking background and ELL students with a Spanish-speaking background.

8. Increased number of subgroups – Research by Thomas Kane (2002) shows a decreasing probability of schools making specific goals as the number of subgroups increases. CRESST has observed that schools with sizeable ELL subgroups tend to have higher numbers of subgroups overall, thereby increasing the probability that that school will not make adequate yearly progress.

9. Possible effect of multiple subgroups – Many ELL students fall into multiple subgroups (*e.g.*, special education, disadvantaged, or a specific racial group). Consequently, under NCLB these student scores frequently count multiple times compared to a student in the general population. Moving these student scores up becomes of critical importance for a school to make adequate yearly progress but may cause a diversion of resources away from other students.

10. Contradiction between NCLB and ELL subgroups – A key NCLB goal is for all subgroups, including ELLs, to reach 100% proficiency in English Language Arts. However, if ELL students were proficient in English Language Arts, they would not be ELL students in the first place. Indeed, if NCLB goals were attained, the ELL subgroup would cease to exist.

NCME Fitness Run/Walk

Run 5K or walk a 2.5K course along San Diego Bay in Embarcadero Marina Park at the Annual Meeting in April, 2004. Commemorative T-shirts will be given to all participants. Check the NCME website in the spring for details.

<http://www.ncme.org>

BALANCE AND SYNERGY IN STUDENT ASSESSMENT

This edition of the newsletter brings two additions to the feature, "Balance and Synergy in Student Assessment," inspired by the National Education Association's (NEA) publication of "Balanced Assessment: The Key to Accountability and Improved Learning." David E. Tanner and Catherine S. Taylor are this issue's contributors.

COORDINATING CLASSROOM AND LARGE-SCALE ASSESSMENT

By David E. Tanner, California State University, Fresno

In the September NCME newsletter, Sue Brookhart quoted a passage from an NEA newsletter: “Under present circumstances there are differences in purpose, intended users, achievement targets, results, roles and responsibilities of those involved, and the manner in which each connects assessment to student motivation.” The question posed for this Newsletter feature, “What should a synergy between large-scale and classroom assessment look like?” is timely. It invites a discussion, not of how the different layers of educational assessment are compatible, nor of how they might be construed so that they appear compatible, so much as how to bring coherence to the assessment whole. The current circumstances, which I believe are described quite accurately in the NEA newsletter, are a recipe for misunderstanding and even conflict, outcomes that are probably inevitable given the different audiences and competing purposes attendant to educational assessment as it is administered in its present forms.

What the classroom teacher wishes to assess may involve quite different learning from what may be at issue in a large-scale state assessment, but need that be so? Is it not possible that both the needs of the classroom teacher to evaluate progress and performance and the need of the state to determine the relative standing of the school or the district could be satisfied by assessment instruments based on the same objectives employing similar kinds of assessment items? In either case, content validity and data reliability should reference the technical quality of the assessment. My experience in teacher preparation reminds me that, at least initially, classroom teachers are fundamentally unconcerned about either calculating or interpreting reliability coefficients. They come by it rightly; neither are many of their instructors. The value of determining the technical quality of assessments is a case yet to be made in many programs. The teachers’ emphasis, after all, is implied in their titles. “Teacher” underscores delivery, not documentation.

As long as classroom and large scale assessments involve distinct objectives there will be a constant tension in the way validity is construed. McColskey and McMunn (Sept. 2003, p. 3) noted that classroom assessments ought not to mimic the format of state assessments but rather focus on “complex performance tasks that assess critical thinking skill in a content area.” This advocacy of some form of authentic assessment is often echoed among classroom teachers. The corollary is that for them the coin of the realm is face validity, an emphasis which often comes at the expense of any attention to data reliability, content validity, or criterion-related validity, all of which tend to be more abstract. When face validity becomes the standard by which large scale assessments are gauged, they are usually found wanting. But ultimately more damaging, a preoccupation with face validity minimizes the relevance of the other more technical qualities which, although less immediately apparent, are probably ultimately more important to the assessment system’s efficacy.

Perhaps the differences in the layers of assessment should reflect only the intended (small, or large) scale. Perhaps rather than discourage efforts to emulate state testing formats at the school and classroom level, there should be an attempt to find a consistency among the different applications in terms of the type of objective, the forms of the test items, and the technical standards by which the assessments are judged.

INCORPORATING CLASSROOM BASED ASSESSMENTS INTO LARGE SCALE ASSESSMENT PROGRAMS

By Catherine S. Taylor, University of Washington

One of the requirements of the reauthorization of the Elementary and Secondary Education Act is that multiple measures be used to determine whether students are meeting state standards. To date, interpretation of this requirement has focused on combining norm-referenced and criterion-referenced test scores, comparing NAEP results with state assessment results, using diverse item types to measure standards (e.g., multiple-choice and performance items), and combining scores from local or classroom assessments with state mandated assessments. While each of these has been discussed in the literature (see for example the summer 2003 edition of *Educational Measurement: Issues and Practice*), discussions have been theoretical rather than practical.

In this article, I discuss practical issues relevant to incorporating classroom-based assessments into large scale assessment programs. I begin by describing the reasons for incorporating classroom assessment information. Next, I discuss the problems that have been identified in the literature. Finally I recommend how to overcome the problems so that the large

scale assessment needs are met while still maintaining local autonomy and teacher/student relevance in the assessment work.

There are three major reasons for incorporating classroom-based assessments: (1) assessing standards that cannot be assessed through paper and pencil tests, (2) increasing the reliability of measurement for standards already assessed through paper and pencil tests (e.g., writing), and (3) providing opportunities for students to demonstrate their proficiency when they cannot perform in a standardized testing situation.

The first reason is fairly self-explanatory. When standards are set, they often include knowledge or skills that cannot be efficiently assessed on a state test. For example, skill in conducting an investigation, recording results, preparing summaries of the data and drawing conclusions, making inferences, and linking results to hypotheses, theory, and concepts are important standards in most states. However, it would be very difficult to assess many of these skills on a large scale test. Richard Shavelson and his colleagues have demonstrated that it takes several hands-on investigations to obtain reliable examinee scores since differences in content influence student scores. Rarely are schools willing to administer multiple standardized investigations in order to achieve the level of reliability needed. Therefore, classroom-based evidence from student conducted investigations would be valuable additions to large scale testing programs. Similar cases can be made for other state standards such as speaking and presenting skills, skill in working with others, and text based research skills.

The second reason for incorporating classroom-based assessments into testing programs is to increase the reliability of scores for knowledge and skills measured by large scale tests. For example, in many states, a single writing prompt is used to assess students' writing skills. It is well known that more than a single complex performance is needed to ensure that decisions made about examinees are reliable. As in science, different writing purposes are differentially difficult for students. For example, students generally find it easier to write narratives than informational pieces. Although states could increase the number of writing pieces, they could also have students collect a variety of different samples of their writing to supplement scores on state tests. This could provide more evidence about students' writing skills across a wider variety of writing purposes.

There are some students who have met standards and demonstrate their proficiency in school every day, yet when faced with a standardized testing situation, they have difficulty performing on the test. This is the third reason for including classroom-based assessments in large scale testing programs. Classroom-based assessments could be used as alternate ways for students to demonstrate their knowledge and skills.

All of these reasons for including classroom-based assessments into large scale testing programs are valid; however, the literature has shown that there are problems with implementing programs that included classroom-based work. Aside from the costs involved in gathering and evaluating collections of evidence from students, three difficulties have been found in the literature: lack of teacher preparation in assessment, inconsistency across different teachers selecting evidence of student proficiency, and insufficiency of evidence in student portfolios. All of these limitations can be dealt with.

For example, lack of teacher preparation in assessment has been documented in the literature for the past 20 years. Yet little has been done to improve the situation. Steps to improve teacher preparation would include required courses in classroom-based assessment during teacher preparation programs (courses truly designed to fit the needs of teachers rather than statisticians and measurement specialists), high quality professional development opportunities for practicing teachers, and model classroom-based assessments that target the standards to be assessed. To date, few states require new teachers to have a classroom-based assessment course prior to initial certification. Where assessment is taught, the topic may be covered in a few weeks of a semester long educational psychology course. In addition, research on classroom-based assessment courses during teacher preparation programs has shown that, when appropriate for teacher needs, these courses are highly valued by the pre-service teachers. However, if the courses are 'tests and measurement' courses, pre-service teachers do not see their value. Therefore, two important steps that can be taken are to increase the required knowledge and skills in assessment for teacher preparation programs and to improve the quality of courses where they are offered. Excellent textbooks are now available for courses in classroom assessment and more are sure to come. Teaching teachers how to create high quality assessments, however, is a labor intensive activity. Teacher education faculty must understand qualities and benefits of high quality assessments themselves, use them in their own teaching, and take the time to guide new teachers as they develop assessment skills. I have found that, once a high quality classroom-based assessment course is available, practicing teachers also want to take the course so they can improve their skills.

The second and third limitations (inconsistency across different teachers in the types of evidence and insufficiency of evidence) can be addressed through clear guidelines for teachers – guidelines that are also accessible to students. Sufficiency of evidence and consistency across teachers requires three types of guidelines. First, teachers and students need lists of acceptable forms of evidence. For example, a writing assessment could be supplemented with a research

report, a written literary analysis, a collection of poetry or a short story, a letter to the editor, and one other written piece that students completed as a normal part of schooling. A science assessment could be supplemented by documentation of the student's performances during lab work, a lab notebook, and a written scientific report similar to those prepared for science journals. Deciding what will be on the list should be part of a public deliberation process. The list should include work that is likely to be an effective way to demonstrate targeted knowledge and skills. The lists should have required entries and may also have choice entries. Research has shown that, when common tools are included in portfolios, the collections are easier to evaluate.

In addition to lists of acceptable forms of evidence, teachers need guidelines for what high quality assessment should look like. Performance task models outline what students must include in their performances. Test blue prints outline what teachers should assess on their tests. For example, a performance task model for scientific inquiry might include the following steps that students must complete:

1. Generate or select a scientific research question related to a local or international issue
2. Conduct background research on what is known about the issue
3. Generate a testable hypothesis about the relationship between two or more variables based on the background research
4. Design an investigation (naturalistic, experimental, model) to test the hypothesis
5. Conduct the investigation using systematic procedures and measurements
6. Systematically record results
7. Organize results in tables, graphs, and/or charts and compute any relevant grade level appropriate statistics
8. Summarize results in own words
9. Relate results to hypothesis
10. Draw conclusions about relationship between results and issue
11. Make recommendations for a) action related to the issue or b) further research
12. Evaluate the success of the investigation in terms of how well it met criteria for investigations

Notice that this guideline gives a message to teachers about what they should be asking students to do in their science classrooms. Suppose a state were to decide to require that students must supplement their science test performances with at least two investigations that included all of the students' written work, photographs of their investigations, and observational checklists (completed by the teacher or peers) regarding systematic procedures, measurement, and recording of results. Teachers would have a very clear message about what they must teach students to do in order that students could complete such tasks on their own. Notice also that the guideline does not require one type of investigation nor does the guideline indicate what topic or concept will be the focus of the investigation. General guidelines make it easier for teachers to integrate these assessments into their instructional plans. If test maps were given as guidelines, teachers would need test maps that help them to focus their tests beyond knowledge. A generic test map for a science test might look like the following:

Type of knowledge	Type of Item			
	Multiple-choice or matching	Short Answer	Essay	Performance Task
Demonstrate knowledge of terms	0-2	0-2		
Describe a processes		2-3	1-2	
Analyze data	1-2			1-2
Apply knowledge to a new situation	1-2	1-2	1-2	1-2
Use rules or formulas	0-2	0-2		
Analyze an example of a concept in a situation			1	1

In addition to test maps, samples of the types of items that would effectively measure the targeted standards would also be helpful. If item types can be boiler plates for other items measuring similar knowledge, conceptual understanding, and skills, teachers are likely to develop better tests that match their instruction. As teachers are more purposeful in developing tests, they may increase the time spent teaching to the important concepts in a subject area.

A third way to enhance consistency of evidence used in statewide assessment programs is to develop rating scales or rubrics that teachers can use to evaluate their students' work. Well developed rubrics or rating scales can help teachers understand what they should be requiring of their students. For example, the following rating scale could help teachers know what they need to ask students to do in their assignments.

Criterion for investigation design	3	2	1	0
Research question can be investigated through scientific inquiry methods			Yes	No
Hypothesis is testable			Yes	No
Design of investigation includes materials and systematic procedures	Thorough	Mostly complete	Attempted	Not done
Design will test the hypothesis		Yes	Partly	No

Guidelines such as test maps, task models, and scoring rules can be created for any assessment of value. Given that most teachers have had little training in assessment, systems be needed to ensure that teachers can work with others to discuss ways to incorporate the performance-based assessments into their regular classroom routine, to share unit plans and instructional strategies, to practice developing high quality items and scoring rules, and to examine collections of students' work so that they can learn how to apply scoring rules. It is possible that, if guidelines such as those described herein were used widely in a state, the quality of teacher assessments would improve along with the quality of evidence used to make decisions about students, schools, and districts.

TEACHER SHOWCASE IN SAN DIEGO

By Judith A. Arter, Assessment Training Institute

Part of the NCME program in San Diego will be a showcase of some 25 San Diego area teachers demonstrating their classroom assessment ideas and practices. Teachers will be invited to showcase their practices in the following areas:

- High quality assessment that results in accurate information on a specific student learning outcome.
- Practices that have been found to improve student achievement and motivation to learn, such as: how learning targets were made clear to students; how descriptive feedback was given to students rather than evaluative feedback (e.g., grades) and how it was used to help students plan their own next steps in learning; how students were involved in their own assessment, tracking progress, and goal setting; and how students were involved in communicating about their own learning over time. This set of teacher practices has been called "assessment FOR learning" by several groups in Canada, Great Britain, New Zealand, Australia, and the US, including the Assessment Reform Group in England and the Assessment Training Institute in the US.
- Effective communication of assessment results to others.

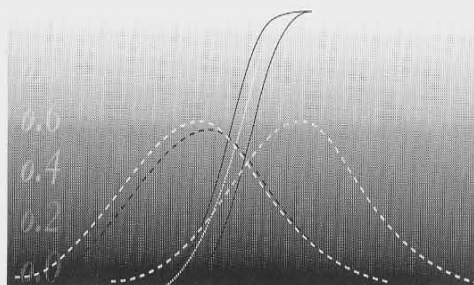
Showcases are an effective way for teachers and measurement experts to see concrete examples of abstract ideas. A previous showcase of Washington State teachers occurred at the NCME 2001 annual meeting in Seattle. Teacher showcases in classroom assessment have been used to great advantage by the Student Aligned Classroom project in Illinois (contact Jay Linksman, jlinksman@pdaonline.org).

We know of one other teacher showcase coming up soon: Cleveland Public School's seventeen "underperforming" schools will present a showcase of classroom assessment ideas as part of the Northeast Regional Professional Development Center's (NRPDC) second annual Data to Information Conference at Cleveland State University on February 11, 2004.

Now Available
on CD-ROM
from the

NCME

national
council on
measurement
in education



JEM

1963–2000

JOURNAL of EDUCATIONAL MEASUREMENT

An Invaluable Tool for the Educational Measurement Professional—

- Capability to search the *Journal of Educational Measurement* by author, by title, or by keyword
- Comprehensive Collection: *JEM* from the first issue through the final issue of 2000
- Articles can be downloaded
- Articles are formatted as they originally appeared

Order today!

Order Form

JOURNAL of EDUCATIONAL MEASUREMENT CD-ROM

Name: _____

Institution: _____

Address: _____

City: _____

State or Province: _____

ZIP: _____ Country: _____

Telephone: _____

E-Mail: _____

JEM CD-ROM

Price in U.S. dollars

Institutional Price \$250.00 x _____ = _____

Individual Member Price \$85.00 x _____ = _____

Domestic Shipping \$3.00 x _____ = _____

International Shipping \$7.00 x _____ = _____

Total Amount Enclosed _____

Payment Method

Check Enclosed ☐

PO Enclosed ☐

Charge My MasterCard ☐

Charge My VISA ☐

Card #: _____

Expiration Date: _____

Daytime Phone: _____

Signature: _____

PO Number: _____

Make checks payable to NCME

Mail order form and payment to

NCME/JEM CD-ROM

1230 17th Street, NW

Washington, DC 20036-3078

**Fax order form and credit card payment
information to**

(202) 775-1824

8/03



**Vice President,
Test Development**

ACT, Inc., a recognized leader for providing assessment and information services for education and business, is seeking a senior executive to manage its Test Development area. In this position the Vice President, Test Development, will manage, direct and coordinate ACT staff in the performance of the following functions in test development for resident and contract programs/services support of ACT's mission:

- Develop assessments for resident and contract programs as assigned.
- Design and develop innovative assessment alternatives that take advantage of technology.
- Participate in the development and implementation of an integrated applied research agenda that supports the design and development of ACT's assessment programs.
- Oversee the design and development of creative assessment approaches that respond to current needs of ACT users.

Successful candidates will have: a doctorate degree in educational measurement, statistics, or related field; previous positions reflecting at least 7-10 years of progressively responsible management responsibilities; a commitment to promoting educational opportunity and human development; strong writing and speaking abilities; and effective interpersonal skills.

ACT offers an attractive compensation package including excellent benefits. For immediate response, email your resume and cover letter, in MS Word or text format, to: Human Resources Department, employment@act.org. For more information, visit our website (www.act.org).

ACT is an Equal Opportunity Employer

**UPDATE FROM THE NCME OUTREACH AND
PARTNERSHIP OPPORTUNITIES**

By Ron Dietel, CRESST/UCLA

The NCME Outreach and Partnerships Committee continues to work to develop productive relationships between NCME and other professional education associations. Among these are the American Association of School Administrators (AASA), the National Association of Secondary School Principals (NASSP), and the National Conference of State Legislators (NCSL). The Outreach and Partnerships committee highly encourages NCME members to submit presentation proposals to any of the three associations. Please contact Ron Dietel, chair of the NCME Outreach and Partnerships Committee (ron@ucla.edu or 310-794-9168) if you are interested in developing a proposal. Brief information is provided below about upcoming conventions and meetings from each association.

American Association of School Administrators

AASA seeks to improve the condition of public education, connect schools and communities, and enhance the quality and effectiveness of school leaders. Although the 2004 AASA conference agenda is finalized, NCME opportunities will be available for the AASA February 17-20, 2005, conference in San Antonio. For a full listing of future AASA conferences see <http://www.aasa.org/conferences/>.

National Association of Secondary School Principals

NASSP promotes the development of administrative leadership and students' intellectual growth. The 2005 conference will be held in San Francisco February 25-28. Proposal information will be available March 1, 2004, on the NASSP web site, www.nasspconvention.org. Proposals must be submitted by May 7, 2004.

National Conference of State Legislators

NCSL is a forum for advancing ideas nationwide and on Capitol Hill. The 2004 annual meeting is scheduled for July 19-23 in Salt Lake City, Utah, and may still have opportunities for NCME members. No Child Left Behind continues to be a major discussion topic amount NCSL members. The 2005 NCSL annual meeting is scheduled for August 14-21, 2005, in Seattle, Washington. NCSL also offers a spring and fall forums on key legislative topics. Contact Scott Young at 303-856-1564 if you have an education topic that you are interested in presenting at any NCSL meeting.