GOOD NEWS FROM THE PRESIDENT
By David A. Frisbie, University of Iowa

This article is my third quarterly report to NCME members regarding the significant organizational changes within NCME that are occurring over the 2004-05 year. I have some very good news to share about what's happening with NCME and how it will impact all of us, both in the near-term and in the long run. Please read on.

Organizational Changes
In June and September, I used this column to describe why NCME had begun looking for a publisher of its journals and why our central office services would no longer be provided by AERA after June, 2005. Here's as update on our progress.

• *Journal Publications***.** I'm pleased to announce that NCME has contracted with Blackwell Publishing Company to publish both of our journals. Blackwell publishes more journals than anyone in the world and has headquarters in both the UK and Boston. Beginning with our 2005 volumes, NCME members and institutional subscribers will receive print copies of the journals as well as on-line access to each issue. Part of our agreement with Blackwell also includes the digitizing of all back issues of the NCME journals so that they also can be viewed and searched on line. (You will learn more soon about *Synergy*, Blackwell's electronic system for searching and accessing information across disciplines world-wide, or go to *http://www.blackwell-synergy.com/*) In addition, our institutional subscribers will be offered our journals as a package, or bundle, rather than as separate subscriptions, a change that will make EM:IP more readily accessible to a wider audience. Finally, our partnership with Blackwell will give increased visibility to NCME and its mission, especially internationally, and will provide us with a more solid financial basis for the future. I'll provide additional details about the Blackwell change at the NCME Breakfast in Montreal.

• *Management Services.* At its October meeting, the NCME Board of Directors chose The Rees Group as the management company to pursue for contracting for central office services. Although this contract has not been finalized at the time of this writing, I anticipate that it will be by the first of the year. The Rees Group is located in Middleton, Wisconsin, a growing urban area attached to Madison. Susan Rees, the president, has been in association management for nearly 20 years and has a staff of dedicated and competent people who provide services to another 15 professional or trade organizations regionally and nationally. Our next Executive Director, Bruce Wheeler, is new to the Rees Group, but he has nearly 15 years of experience in association management. The fresh look and new ideas they'll bring to us provide a terrific opportunity for us to revisit our core mission, operational procedures, and organizational governance. I'll offer more details about the change to the Rees Group in the March newsletter and during the NCME Breakfast in Montreal.

• *Annual Meeting.* Now that we have identified a management company to provide central office services for NCME, we will more actively pursue a contract with AERA to provide certain services that will permit us to continue a joint annual meeting with AERA. The NCME Board believes it is in the best interest of our members to continue this relationship, and as indicated in my exchange of letters with AERA President Marilyn Cochran-Smith, the AERA Council has affirmed its desire to continue the joint meeting with NCME as well. We will work with the AERA staff and the Rees Group staff to have a contract in place by this spring so that the coordination, planning, and implementation associated with the 2006 annual meeting in San Francisco will continue.

Other Issues
In view of the changes noted above, it is more important than ever that each member use the *NCME website* (http://www.ncme.org/) for current information about organizational changes as well as on-going activities. Among the changes you can expect to see in the website within the next year is a "members only" section, which can be used for such things as updating your contact information, paying dues, and accessing our journals (and Blackwell's *Synergy*) on line.

By now, individuals who submitted proposals for our *annual meeting* in Montreal have been notified about acceptance, and you have seen many of the program highlights as listed in the fall issue of EM:IP. We expect the rooms in the Sofitel Hotel that will be used for NCME sessions to be in close proximity to one another. Thus, the opportunities for social and professional networking should be enhanced. I encourage you to monitor the NCME website for information on international travel and any last minute changes in regulations that might affect your safe and uneventful travel to and from Montreal.

As detailed in the fall issue of EM:IP, the *Training* and Professional Development Committee has identified more than a dozen sessions that are scheduled to be offered on Sunday or Monday preceding our annual meeting. Check the NCME website when you register for the annual meeting to get details about training sessions and to register for them on-line. It is particularly important to register for training sessions well before you arrive in Montreal: some sessions will be cancelled if enrollment is too small and others will be filled due to enrollment caps required by room capacities.

I mentioned in the last issue of the Newsletter that the *Membership* Committee has begun reconciling the membership database provided to it by the AERA central office. In particular, the committee is targeting individuals who, according to our records, have lapsed memberships prior to January, 2004. (A mailing will be sent to those whose membership expired during 2004 but who have not yet renewed.) As part of this effort, the committee will depend on an individual in each of the larger organizations in which there are multiple NCME members to contact colleagues with lapsed memberships. We want to increase our numbers but also determine why the memberships are not renewed. I hope you'll be able to help if you're contacted.

At the NCME Breakfast in Montreal, there will be an opportunity for members, as there is each year, to complete a questionnaire that solicits advice about future programming as well as member interest in serving NCME in any of a variety of ways. With the changes I've described here, you can see that this is both an interesting and meaningful time to be involved in NCME. Please take time to respond and to offer your service. If you are unable to attend the annual meeting Breakfast, please go to the website homepage and find a copy of the questionnaire under "Annual Meeting."

PROCEEDINGS FROM A RESEARCH CATALYST CONFERENCE HELD BY THE NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS TO INVESTIGATE THE IMPACT OF STANDARDS

By Carol S. Parke, Duquesne University

In the fall of 2003, the National Council of Teachers of Mathematics (NCTM) held a research catalyst conference at their headquarters in Reston, VA. The purpose of the conference was to frame research questions and catalyze coordinated research on the impact of national and state mathematics standards on policy, assessment, student achievement, teaching practices, curriculum, and instructional materials. The content-based standards movement in mathematics education began more than a decade ago when NCTM released the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). Several related *Standards* documents followed, and an updated version of the *Standards* was published in 2000. Although many studies related to the mathematics standards have been conducted over the years, they have been largely anecdotal in nature. NCTM recognized a need for more systematic and rigorous research to better understand the role and influence of standards. Given the implications of the No Child Left Behind legislation, the need for such research had become even more essential. NCTM was also interested in creating an interdisciplinary research community in which a professional cadre of scholars came together to discuss research agendas. Thus, the research catalyst conference was planned and enacted.

A select group of approximately 100 researchers, policymakers, mathematicians, and mathematics educators were invited to attend in a three-day conference, during which time they participated in one of eight working groups. Each group focused on a specific topic, and the activities and discussions throughout the sessions were facilitated by co-chairs. The topics of the eight groups were: assessment and student achievement, state and national policy, local policy and community context, changing nature of schooling and school demographics, teaching and learning, instructional materials and curriculum, teacher preparation, and teacher quality and professional development.

Outcomes of several working groups may be of interest to the educational measurement community. For instance, the Assessment and Student Achievement working group developed a framework for organizing and defining research on the influence of *Standards* on assessment. The framework includes several dimensions such as the form and function of the assessment, the level of the assessment (international, national, state, district, school, and classroom), and the stakeholders (policy makers, administrators, practitioners, students, parents, public). The State and National Policy working group defined several areas of inquiry, such as investigating the direct impact of the *Standards* on state and national policy and studying the alignment among federal, state, and local policies regarding standards. Discussions in the Local Policy and Community Context working group revolved around clarifying the structures and mediating processes that may influence implementation of the standards at the school level and teacher level. Finally, a set of research questions framed by participants in the Instructional Materials and Curriculum working group focused on the alignment between state curriculum, state assessment frameworks, and instructional materials used by teachers.

Members of NCME may also be interested in panelist presentations at three plenary sessions held during the conference. Current issues and trends in educational research was the focus of the first session. Panelists in the second session represented national, state, and local policy as well as the fields of mathematics and statistics. They discussed the impact of standards from their own experiences and perspectives. The final session included responses and observations from international colleagues and a representative of the National Science Foundation.

The discussions and outcomes of each of the eight working groups are documented in the *Proceedings of the NCTM Research Catalyst Conference* (NCTM, 2004). Presentations at the plenary sessions as well as eight research briefs are also contained in the *Proceedings*.

National Council of Teachers of Mathematics, NCTM: 2004, *Proceedings of the NCTM Research Catalyst Conference*, Reston, VA. (visit http://www.nctm.org for information on obtaining a copy of this document)

National Council of Teachers of Mathematics, NCTM: 2000, *Principles and Standards for School Mathematics*, Reston, VA.

National Council of Teachers of Mathematics, NCTM: 1989, *Curriculum and Evaluation Standards for School Mathematics,* Reston, VA.

JOINT COMMITTEE ON STANDARDS – REQUEST FOR PARTICIPATION
By Donald Yarbrough, University of Iowa

The Joint Committee on Standards for Educational Evaluation (JCSEE) is currently revising its *Personnel Evaluation Standards* and its *Program Evaluation Standards*. NCME has a long history of contributing to these practice standards and sponsors a representative on the JCSEE, as do 15 other national professional organizations. This is the first revision in more than 15 years for the *Personnel Evaluation Standards* and more than 10 years for the *Program Evaluation Standards*. NCME members contributed substantially to the development of both sets of standards, and your contributions are crucial to ensure the quality of these revisions. Below are letters to the membership of NCME from the Task Force Chairs for each of these two revisions. Please read the letters and volunteer as appropriate.

From Barbara Howard, Chair, *Personnel Evaluation Standards* Revision Task Force.

Dear Colleagues:

As experienced professionals, you no doubt realize that engaging in sound practices of personnel evaluation can result in more effective educators and higher quality education. The Joint Committee on Standards for Educational Evaluation (JCSEE) issued the first *Personnel Evaluation Standards* in 1988. While still useful in the field for guiding personnel evaluations, we are now revisiting these standards to revise them and issue the second edition. We have just completed the first draft of this revision.

Please help us ensure the high quality of these materials. As a person contributing to education and involved in its professional organizations, we invite your help in reviewing and improving this draft material. Please provide your comments and suggestions on one of the four attributes of sound personnel evaluation practice: Propriety, Utility, Feasibility, or Accuracy.

All reviews may be completed and submitted electronically through a secure Web site. Your review assistance will be acknowledged in the final document published by Corwin Press, a division of Sage Publications. We also are seeking sites for field trials of these standards during the spring of 2005. Field trials are opportunities to try out the materials as part of your everyday work and report back on their usefulness and quality.

To accept this invitation to review and/or participate in a field trial, please e-mail Dr. Barbara B. Howard, Task Force chair, SERVE Regional Educational Laboratory, bhoward@serve.org To learn more about the Joint Committee's standards development work and its standards for educational evaluation, please visit this Web site: <www.wmich.edu/evalctr/jc/>.

From Don Yarbrough, Chair, *Program Evaluation Standards* Revision Task Force.

Dear Colleagues:

Many of you have already responded to the Joint Committee on Standards for Educational Evaluation (JCSEE) survey regarding revisions to the *Program Evaluation Standards* (the third edition is tentatively schedule for completion in 2006). However, because we want the broadest possible input to the *PgES* revision process, we are making the survey available from now until December 31, 2004 for those who have not yet responded. If you have not yet filled out the survey, please log on at http://ateserver.cs.wmich.edu/PgES2004

You will be asked first to create you own unique user login and ID. Then you must log in with that ID in the box just above the box where you created your unique login. From that point on, you can come back to the survey as many times as you want to work on it until you finish and "close it out." Each time you come back, you should log in with that same unique ID you created. If you experience problems with the above site or prefer to fill out a paper survey, please print off a copy of the following PDF file (http://projects.education.uiowa.edu/pges2004) and mail it to the address at the bottom of this invitation (and on the paper copy). If you experience difficulty with that procedure, you may contact me directly, and I will mail you a paper survey.

Please fill out those portions of the survey that are pertinent to you (this can take from 5 minutes to much longer, depending on the length of your comments). The survey provides you with ample room for comments and asks if you want to volunteer to review the draft revision or serve as a field tester. If you prefer to volunteer or to send your suggestions and comments directly, you may do so at anytime by emailing me at d-yarbrough@uiowa.edu. In addition, if you want to provide more detailed comments, suggestions, documents, articles, references, or other information, you may send these directly to me:
Donald B. Yarbrough, *PgES* Task Force Chair
210 Lindquist Center South
The University of Iowa
Iowa City, IA 52242

You can also send your comments about the process the JCSEE is following (American National Standards Institute procedures). For additional information about the *Program Evaluation Standards* and this revision process or about the JCSEE and its sponsors, members, procedures and background, please see the following Web site: http://www.wmich.edu/evalctr/jc/

Lastly, because many of you are members of several national organizations that sponsor the JCSEE, you may receive multiple copies of this invitation. You only need to fill out the survey one time. When you fill out the survey, you can list all the organizations to which you belong. Thank you in advance for your assistance with this important task. If you have not yet made your wishes known about revisions to the *PgES,* we sincerely invite you to do so now.

NCME ANNUAL MEETING

MONTREAL

APRIL 12-14, 2005

watch for more information:

www.ncme.org

ASSESSMENT AND A CULTURE OF CONFIDENCE
By Rick Stiggins, Assessment Training Institute

Successful classroom assessment has both a technical side and a human, emotional side. To promote student success at learning, teachers must assess accurately and manage assessment data efficiently. That's the technical side. But a far more important key to using assessment to help students succeed is to assure a positive emotional reaction to classroom assessment results by the student. If assessment results are to be beneficial (that is, are to promote continued learning), they must help the learner believe academic success is within reach for them if they keep trying. Results, and how they are used and communicated to students, must help students remain confident. Confidence breeds optimism—the expectation of a positive learning results in the future. Such expectations help the learner remain willing to take the risk of trying to learn.

To explore these emotional dynamics of classroom assessment in a bit more detail, construct a two-by-two chart on a sheet of paper. Label the top row "students who succeed at learning," and the bottom, "students who fail to learn." Label the left column "assessment says student succeeded," and the right, "assessment says student failed."

So the upper left cell contains students who learned and the assessment verified that fact. In the upper right, we find students who learned but the assessment says they did not. The lower left includes those students who failed to learn, a fact missed by the erroneous assessment. And the lower right cell says the student failed to learn and the assessment reflected that result.

Now consider the student's likely emotional reaction to the assessment results in each cell. That reaction will drive what the student does in response to the result—that is, whether the consequence of assessment is productive or counterproductive in terms of that student's ultimate success in future learning.

The student in the upper left cell is affirmed. It feels good to be a winner. The student aspires to keeping the wins coming. Optimism will reign, bringing an expectation of positive results in the future. The focus for this student will be on what he or she does well. This student is likely to put forth effort in the future and to continue to risk trying to learn even in the face of challenges because of this record of success.

In the upper right cell we find the successful student inappropriately judged to have failed. This will produce a different emotional reaction. Inept assessment is likely to result in confusion and defeat, because the student typically doesn't know that the assessment was wrong. Initial panic will be followed by frustration, and then resignation, if the mistake goes uncorrected. If the error stands, a kind of cynicism will result that reduces desire to learn, risk taking and thus future success. Why even try if this is going to happen? Obviously, this is a counterproductive result.

These same emotions can overtake the student in the lower right—the failing student in fact and in assessment results. The danger here is that the pattern of failure may become chronic and thus appear to be inevitable in the mind of this student. Pessimism will come to reign, bringing an expectation of a negative result in the future. This hopelessness and feeling of vulnerability feeds the self-fulfilling prophecy. This will shorten the amount of effort the student is willing to expend before giving up. Self criticism will become easy given the previous record. The impact on later learning can be devastating. Again, we see a counterproductive impact.

Our final cell is the lower left, where the student appears to have beaten the system, perhaps unwittingly. If the student doesn't know the assessment results are incorrect, assumes preparedness to move on and does so, the lack of prerequisites may doom that student to later failure. If the student knows the results to be incorrect, then cynicism can emerge about the teacher's ability to detect what's really happening in the classroom. This is not likely to spur greater effort in the future.

So in three of the four cells, we uncover the risk of emotional reactions that are counterproductive to learning, either due to inaccurate assessment or accurate assessments that leave the learner without hope of success in the future. This is why we encourage teachers always to (1) gather dependable information via classroom assessment and (2) use the classroom assessment process and results to keep students believing that success is within reach if they keep trying.

If classroom assessments are to be accurate, we must help practitioners tune in to the different information needs of different assessment users, articulate clear and appropriate classroom-level achievement targets, select proper assessment methods and develop them well, and communicate results effectively to intended users. To use classroom assessment to maintain student confidence, research indicates that teachers must involve their students in the ongoing classroom assessment, record keeping and communication process, so as to help them understand that success is within reach if they keep trying.

SUMMARY OF TEACHER AND ADMINISTRATOR LICENSING STUDY
By Mike Herrick, Herrick Research

The Assessment Training Institute Foundation, Inc. of Portland, OR contracted with Herrick Research of Eau Claire, WI in June, 2004 for the purpose of conducting an investigation of teacher and administrator licensing requirements in student assessment nationwide. The Licensing and Certification Units in the departments of education in each of the 50 states plus Washington D.C. were contacted in a variety of ways: telephoning state Licensing Coordinators, distributing e-mail questionnaire to Licensing Coordinators, checking state licensing and certification websites, and checking manuals and publications about state licensure requirements.

Teacher Licensing. The study revealed no evidence that any of the 51 states explicitly require a course in student assessment as part of their state's teacher preparation program. Rather, the study revealed a move away from specifying course requirements for teacher licensing and towards the specification of professional standards for teachers. Thirty-nine states reported that issues related to assessment, testing, and/or measurement are integrated into basic required courses dealing with the knowledge of teaching.

Nearly all states have published standards of teaching competence. A total of 49 states have either written their own standards or adopted standards from the Interstate New Teacher Assessment and Support Consortium (INTASC)—35 states are members—or from the National Council of Accreditation of Teacher Education (NCATE)—10 states are members—or both. The one INTASC standard that deals with student assessment reads, "The teacher understands and uses formal and informal assessment strategies to evaluate and ensure the continuous intellectual, social and physical development of the learner." Under this standard are listed specific knowledge (strengths and limitations of various assessment methods; how to select, construct, and use various assessment methods; validity, reliability, bias, and scoring), dispositions (value assessment as an essential part of instruction), and skills (use assessment for tracking progress, planning instruction, and student-involvement).

A total of 45 states require some type of exam to become initially, provisionally, or professionally licensed. Twenty states require the Praxis I exam, an ETS product that measures reading and math. Eighteen states require the Praxis II exam, an ETS product that measures content knowledge. A total of 22 states reported that they require some kind of performance assessment for the initial and/or the professional license. That performance assessment may take various forms such as the formal ETS Praxis III performance assessments, writing assessments, portfolios, or competency ratings from student teaching. No direct measures of assessment competence are required.

A total of 38 states reported that competent teaching performance must be demonstrated in some way before an initial or provisional teaching license can be granted and/or a permanent or standard license can be granted. Of those 38 states, 24 reported that teacher competency is determined by the teacher training institution prior to program completion for an initial license, typically by student teaching evaluations, and thereby not determined by the state. States gave no indication that assessment performance is directly considered for teacher certification.

Administrative Licensing. As with teacher licensure, there is no evidence that a course explicitly devoted to student assessment is required for administrative licensure in any of the 51 states. Very few states reported specific courses of any kind that need to be completed for an administrative (school or district administrator) license. A total of 31 states require candidates for school administration licenses to pass some sort of a state specified examination. Of these 31 states, 13 require that candidates pass the School Leaders Licensure Assessment (SLLA). The SLLA is a set of performance- based assessments developed in collaboration between the Interstate School Leaders Licensure Consortium (ISLLC) and ETS. No direct measures of assessment competence are required.

Twenty-two states reported having standards defining competence for school administrators. Of these 22 states, 17 belong to the Interstate School Leaders Licensure Consortium (ISLLC) and can be presumed to model their standards from ISLLC's six standards. None of them deal specifically with student assessment. Only 4 states were found to have administrator standards regarding assessment.

Conclusions. Perhaps the most significant finding in this investigation is the fact that no college or university course in student assessment is required for teacher or administrator licensing, and there is no evidence that states actually require assessment competency for certification. The primary conclusions are: (a) requirements for assessment courses have been replaced by teacher and administrator standards; (b) there has been a rise in the role of INTASC and ISLLC in standards development for teacher and administrator licensing; and (c) formal testing has dominated the measurement of proficiency of teachers and administrators for licensing and certification. However, there is no evidence that these tests include domains directly related to competence in student assessment.

ASSESSING INSTRUCTIONAL AND ASSESSMENT PRACTICE
By Anne Lewis, Consultant

The following is an excerpt from the forthcoming proceedings of the 2004 CRESST Conference, Research Guidance: Assessment, Accountability, Action! Full proceedings will be available on the CRESST web site by June 2005. Powerpoint slides from each presentation are available at www.CRESST.org.

In the symposium, *Assessing Instructional and Assessment Practice*, researchers Brian Stecher, Hilda Borko, and Noreen Webb presented findings from three different studies on the evaluation of classroom instructional and assessment practices. In the last session, Dylan Wiliam described one possibility for a more effective learning environment, including formative lessons and assessments.

Using Classroom Vignettes to Measure Teachers' Instructional Practice in Mathematics
Brian Stecher, Vi-Nhuan Le, and Laura Hamilton, RAND Corporation

Studies of large-scale educational programs often need accurate descriptions of classroom practices. However, such descriptions are difficult to obtain in an efficient and timely manner. Common methods for measuring instructional practice have limits that can affect their validity. Observations can be complex, time consuming, expensive, and subject to the biases of the observer. Similarly, survey responses can have a self-reporting bias, be distorted by faulty memory, and suffer from the lack of a shared understanding of reform terminology. Maintaining classroom logs has some of the same problems.

To address this problem, Stecher and his colleagues developed a short "vignette-based" measure of mathematics instructional practices that asks teachers to rate the degree to which various teaching practices correspond to what they do in their own classrooms. The Vignette-Based Study of Reform Teaching Practice, part of the Mosaic II Project funded by the National Science Foundation, tries to measure intention to engage in reform-oriented teaching. Teachers in the project respond to specific, hypothetical, but familiar situations with alternative stated in clear behavioral terms. This approach presents realistic situations and choices, using common terminology, which standardizes the collection of teacher responses.

In the study, a panel of math experts used the National Council of Teachers of Mathematics standards and other documents to create vignettes for two common 4th grade math topics. The options given to teachers represent a range of high- and low-reform actions and are parallel across the two topics. Each option is assigned a reform value, which allows the set of responses to be combined into an overall measure. An example is below.

After praising both groups for using effective strategies, how likely are you to do each of the following in response to these two explanations?

(Circle One Response in Each Row)

		Very unlikely	Somewhat unlikely	Somewhat likely	Very likely
a.	Ask the class if they can think of another way to solve the problem	1	2	3	4
b.	Suggest that the class check the results by using a calculator				
c.	Tell them the first group's method is faster	1	2	3	4
d.	Tell them they are both right and move on to the next problem				
e.	Have a classroom discussion about the differences between the two approaches	1	2	3	4

Analyses with 80 fourth-grade teachers showed consistent responses across the two vignettes, and responses were moderately correlated with those obtained from classroom observations and more traditional teacher surveys and logs. The results provide guidance to inform the development of measures of instructional practice.

The researchers learned that the structured vignettes were difficult to develop, reading demands were high, and the evidence for validity was mixed. For example, the study found that teachers' responses were stable across parallel math contexts. The "reform inclination" scale derived from the vignettes correlated with several survey and log measures of reform practice, but not with observations. However, the "Euclidean scale" of reform, which was derived in terms of the distance of each teacher from an ideal high-reform teacher, correlated with observational measures of reform-oriented practice, but not with surveys or logs.

The next step is to study the relationships of vignette-based measures with student outcomes, particularly achievement, which is the overall purpose of the Mosaic Project. The quality of the vignettes needs to be improved through interviews with teachers, the effect of the length and level of detail in the vignettes on teacher responses needs to be evaluated, and the project needs to explore how the vignettes could be used in other contexts. Studies are needed to determine if the vignette approach can be brought to scale.

Using Classroom Artifacts to Measure Instructional Practice in Middle School Mathematics: A Two-State Field Test

Hilda Borko, Suzanne Arnold, Beth Dorman, Karin Kuffner, Colorado University-Boulder;
Brian Stecher, Mary Lou Gilbert, Alice Wood, RAND Corporation

Another experiment for documenting teacher practice is to collect artifact packages, or "Scoop Notebooks." The notebooks scoop up a typical week's worth of instructional materials: lesson plans, assignments, tests, student work, photographs of the classroom, teacher reflections on class sessions, and student work. The basic question that frames the collection of materials is, "What it is like to learn math in your classrooms?" The research project is based on the premise that an artifact collection such as the Scoop Notebook has the potential to overcome limitations of surveys and case studies as methods for measuring instructional practice by representing what teachers actually do in their classrooms rather than what they report that they do, while requiring fewer resources than case studies.

The researchers used data from 30 middle school teachers in California and Colorado to analyze the reliability and validity of the Scoop Notebook. Data included notebooks completed by the teachers, researcher observations, and audiotapes of lessons (for a subset of 8 teachers). They developed a scoring guide on 11 dimensions of instructional practice, such as cognitive depth, problem solving, and assessment, which raters used to score the notebooks, observations, and transcripts of audiotaped lessons. The researchers found moderately high to high levels of agreement among raters along the 11 dimensions, for ratings based only on the Scoop Notebook.

One set of validity analyses compared ratings based only on the Scoop Notebook with ratings based on the Notebook plus another source of data (observations or transcripts). The researchers found moderately high levels of agreement for all dimensions, although agreement was lower for some dimensions such as Mathematical Discourse and Assessment. "Some dimensions and teaching practices present greater challenges than others for artifact-based tools such as the Scoop Notebook," Borko said. In addition, teachers vary their activities from one day to the next, and "raters don't always agree on the weight to be given to different activities." Disagreements among raters may be greater when there are inconsistencies in the data due to variations in a teacher's instructional practices.

In another set of validity analyses, the researchers found a substantial difference between ratings of Colorado teachers, who were using more reform strategies, and California teachers, who used more traditional strategies (the former had higher scores). Borko concluded that the artifact collection "is useful for describing classroom practice in broad terms, but it should not be used to make high-stakes decisions about individual teachers." She recommended more research on the differences in scoring among the raters and explorations of why some classrooms and some dimensions are more difficult to rate than others.

The Effects of Teacher Discourse on Collaborative Student Problem Solving
Noreen Webb, Marsha Ing, Nicole Kersting, Kariane Nemer - CRESST

This study attempted to find out why students use such low-level discussions in collaborative groups, even when they have had extensive preparation and practice in collaborative problem solving. For this study, researchers examined teacher modeling of discussions. In four 7th grade general math classes, students worked in heterogeneous groups for one semester. All teachers and groups were audiotaped for five class periods, and the students were given pre- and post-tests.

The students and teachers received ongoing training for group work. Students who understood the problem, for example, were told to refer to a "helper" classroom chart with such guidance as: "Notice when other students need help." "Tell other students to ask you if they need help." "Be a good listener." "Give explanations instead of the answer." "Check for understanding." "Praise your teammates." Students who did not understand the problem could refer to a "helper" classroom chart that told them to: "Recognize that you need help." "Choose someone to help you." "Ask clear and precise questions." "Keep asking until you understand."

The researchers found that despite the training on higher-level group work, teachers did not noticeably change their style of instruction. In most cases, the teacher presented the steps in a problem, requested numerical answers only, did not explain why student answers were correct or incorrect, did not probe student thinking, and focused exclusively on numerical procedures. This was not surprising, Webb said, "because most of the training was for students." While teachers practiced the same training activities, teachers focused on how students should behave during small group work rather than how teachers might use the principles of effective helping in the training activities to inform and change their own classroom instruction."

Despite the wall charts and training to support high-level helping behavior, the students usually followed the teachers' modeling and peer assistance was generally poor, *e.g.*, helpers dictated the calculations. Helpers didn't try to determine a help-seeking student's level of understanding either before or after providing help. Students seeking help did not ask specific questions or reveal their level of understanding, nor did they use the help to test their understanding.

The researchers found few instances of "either teachers or students trying to find out the thinking of students needing help…, and it was rare for these students to explain what they were having trouble with, or to use the help to try to solve problems on their own." Basically, Webb concluded, the students "did what the teachers did, not what the teachers said." A next step for the research is to identify teachers with useful helping styles and determine if students adapt that same helping style when assisting other students.

Assessing Instructional and Assessment Practice: What Makes a Lesson Formative?
Dylan Wiliam, Educational Testing Service

One of the stumbling blocks to studying instructional and assessment practice, according to Dylan Wiliam of the Educational Testing Service, is that "real lessons do not use high-quality instruction." He reported on research at ETS and formerly at Kings College in England, that defined and searched for formative lessons and assessments that actually shape learning. The role of teachers, he noted, "is not to teach, but to create environments in which students learn. Right now, teachers are working too hard and students, not enough." High quality formative lessons and assessment provide students strong feedback. In the formative context, assessments should monitor, should be diagnostic and should move from "what is wrong to what to do about it." The first priority of formative assessments (which can be external) is to serve learning, he emphasized, not accountability.

Wiliam showed the following chart to help describe good formative assessment and the role of teachers and students.

	Where the learner is	**Where the learner is going**	**How to get the learner there**
Teacher	Evoking information	Curriculum philosophy	Feedback
Peer	Peer-assessment	Sharing criteria	Peer-tutoring
Learner	Self-assessment	Sharing criteria	Self-directed learning

In the above table, for example, a learner might use self-assessment to establish their current knowledge, and as a result of activities provided by the teacher, be clear about what it is they are trying to achieve. If they then know what to do, they would be able to engage in self-directed learning, although more typically they might get support from peers, or the teacher.

Formative assessment can be thought of as an aspect of the regulation of learning. A key aspect of the regulation of learning is the creation and use of "moments of contingency," those times in a lesson where a teacher can go in different ways, depending on the evidence the teacher has about the extent of the students' learning. There is increasing evidence that the use of formative assessment increases student achievement, even when such achievement is measured via state-mandated standardized tests.