Most of us believe we understand what we mean by performance standards—it’s part of the other sandwich (apologies to Governor Roy Romer) between content standards and performance assessments. As the litany goes, content standards tell us what, performance standards tell us how well, and performance assessments operationalize and measure the whole thing.

There is little agreement on what performance standards are, how they are best set, and what their relationship is to details used in scoring student performance.

As we approach the implementation of standards-based assessments either for reporting our progress toward the National Education Goals, for allocating resources, or, in some states, for certifying students, it becomes clear that our sense of the obvious and our global understandings of performance standards get shaken by the realities of making these ideas work in a real system. The contention about the appropriateness and meaning of the achievement levels of the National Assessment of Educational Progress (NAEP) is simply one more set of evidence that we have not completely thought through the problem.

In fact, there is little agreement on what performance standards are, how they are best set (other than using particular jury-type procedures) and what their relationship is or should be to details used in scoring student performances.

Two different traditions give rise to the idea of performance standards for student work in the U.S. The first, deriving from instructional systems development, most widely used in the military, focuses on performance standards as embodying statements of conditions, behaviors, content, and levels of attainment in a particular situation, and the desired attributes for particular go/no-go decisions. The attainment levels could be qualitative (as exhibited in either criteria for constructed performance or actual models of constructed performance)
Towards an Understanding of Performance Standards...from page 1

or quantitative (expressed as percentages of achievement on criterion-referenced tests).

A second approach is the familiar setting of a “cut-score” typically based on achievement data on a unidimensional scale. Here the assumption is that a higher score reflects better achievement because item difficulty varies. For these cut-scores to work as performance standards, against which educational programs may be improved, we must know a good deal about how students’ cumulative scores correspond to the actual items students answer correctly.

Clearly, the inception of performance standards in the context of national standards requires a melding of these two approaches: the identification of attributes of performance that can be achieved in a standards-referenced context and sorting out what different levels of attainment mean psychometrically.

Levels of Performance

One immediate problem is to decide on how many levels of performance we are going to report against and what we are going to call them. NAEP chose three: Basic, Proficient, and Advanced. Some states have four, five, or six. In all there are an assumption that performance is cumulative: People at level five also can do levels four, three, two, and one. Although the rhetoric supports the idea that higher levels call for qualitatively different skills and understanding than lower levels, we know such relationships won’t be perfect (we can all remember missing a really simple test item after doing the most complex problem perfectly). Still, we must carefully match descriptions of performance standards to expected differences in tasks attempted and attained.

The aggregation of standards for individual tasks to establish standards for overall performance on a collection of tasks can be readily accomplished by applying mechanical rules. But, justifying the use of a particular rule remains problematic. Furthermore, aggregation of task-by-task judgments may lead to substantially different overall standards than judgments of complete records of performance for individual students.

We are going to have to find ways to acknowledge in our systems qualitatively different performance.

Scoring Performance

A number of different approaches are possible for scoring performance. Some developers are creating assessments with rubrics that directly correspond to achievement levels: If a state wishes to report out at five levels, then the rubric will have score points for each scored dimension. That approach may be efficient, but it is also very possible that certain tasks will not lend themselves to such a direct correspondence. For example, we can imagine that there is a number of different ways to be “satisfactory” in an essay, and neither meeting a level “3” nor averaging scores will likely be an adequate approach for some tasks.

We are going to have to find ways to acknowledge in our systems qualitatively different performance, particularly those patterns of accomplishment that aren’t nice and neat and cumulative. Part of the problem may be in the routinized way we plan to go about reporting these new performance-based data. If we squeeze new wine into old bottles, some of the reason for performance assessment in the first place may be forgotten.

CREST researchers are attempting to address this problem through their work with the Technical Review Panel of NAEP, through their work with the Council of Chief State School Officers and with individual states working on performance levels, in innovative work with school districts, and in the international arena, attempting to learn what we can about what elements might or definitely should not be imported from our colleagues beyond our borders. Help us by identifying key problems you have in the setting of performance standards.

Alternative Assessments in Practice Database

Containing listings from over 250 developers of new assessments, the Alternative Assessments in Practice Database will be of special use to teachers, school district administrators, assessment developers and others interested in new methods for assessing student growth. The database contains detailed information about each assessment, including subject matter and skills measured, assessment type and purpose, scoring characteristics, and availability of the assessment.

The assessments focus on a wide variety of subjects ranging from the traditional to the less common. Many of the new assessments contain short-answer items, essay questions, demonstrations, use of manipulatives, hands-on tasks, non standard written products, and experiments. Although the assessments themselves are not included in the program, many developers have indicated a willingness to share samples or prototypes.

Inexpensively priced at $15, the easy-to-use program includes a single double-density disk, a 50-page manual with comprehensive, step-by-step illustrations, and a built-in “on-line” help feature. As a bonus, the manual contains a chapter called “Introduction to Alternative Assessment.” To order, call Kim Hurst at (310) 206-1532.
Real Change Is Real Hard: Lessons Learned in Rochester
by Adam Urbanski

So if we cannot have standards that are achievable by all students, or at least most students, then I would raise the question of whether or not they’re worth having. Because excellence without equity is not excellence.

Adam Urbanski is president of the Rochester (NY) Teachers Association and vice-president of the American Federation of Teachers. The following remarks were made during his presentation at the 1993 CRESST Conference “Assessment Questions: Equity Answers.”

When you are the final speaker at a conference you should be mindful of the three cardinal rules of effective public speaking: be brief, be sincere, and be seated. So let me briefly and sincerely share with you some thoughts related to real classroom change in the context of what we have learned in Rochester.

I respectfully suggest that all students are learning already. They’re just not learning the same things. Some students are learning math and English and foreign languages and the arts and physics. And some are learning a lot about exclusion and failure and discrimination and lack of opportunities and low expectations of them. But you can’t stop students from learning. You can only assist in channeling them to certain “kinds” of learnings. That is why the whole issue of standards and assessment is so germane. They are absolutely essential and not only essential, but indeed the necessary starting point for all other reforms.

School change need not be a choice between making things better or making things fair. We’re capable of doing both. We must make things better and be fair in doing it. We must not exclude some students from high standards because in addition to exclusion, the most insidious form of racism is to lower standards for any group for reasons that are not germane. And so, rigorous standards for all kids is the right starting point for our half of the reform agenda.

But when it comes to fixing the lot of kids, I think educators own only half of the agenda. We educators have the responsibility to make schools more ready for all kids. Schools are not now ready for all kids. And the flip side of that, the other half of the agenda, is for each community, and society in general, to make all kids more ready to learn. One without the other will not work. Schools eminently prepared for children to learn, but full of children who are not ready to learn, will not be effective. And schools full of children ready to learn, but not equipped to help them learn, will not be effective.

Without standards, education reform will continue to be rudderless.

Standards doesn’t have to mean making things more rigid. Standards ought to mean setting benchmarks for what quality is. Or, as Grant Wiggins would put it, for what is worth doing and for what is worth measuring. Without standards in education reform—and I can tell you from experience, having been involved in this over the last decade, and having the scars and lumps to prove it—without a clear consensus about what students should know and be able to do, education reform will continue to be rudderless, and we will likely have education reform anarchy: Anything goes—your notion is as good as mine or anyone else’s because there are no benchmarks.

But standards, as critical as they are in my view, are only one of a number of necessary ingredients to help make schools effective. So standards are necessary but not sufficient to make schools effective. We will not have achieving schools just because we get standards right. We will not have achieving schools just because we get standards and assessments right. What we will achieve is merely getting standards and assessments right. This alone doesn’t mean that kids will learn and teachers will teach and schools will function appropriately. I would offer—this is no correlation to the fact that the Pope is Polish—ten principles for achieving schools. Let me share them with you.

Consensus
First and foremost, we must have a consensus about what students should know and be able to do. If we do not have that, we have no right to know what adults should know and be able to do. And we have no basis for knowing how schools should be structured and organized. Because without consensus about
standards, we don’t know what the primary purpose of it all is.

Small Schools
In addition to standards, what we need for achieving schools, and what will otherwise hamper achieving those standards, are small schools, small enough so that students are well-known.

Active Learning
Third, active, real-to-life learning rather than the drill-and-kill passive by-listening-rote schooling that now inhibits learning and makes it virtually impossible to achieve any worthy standards.

Authenticity
Fourth, authentic assessments—that is, if the curriculum is active so should be the measuring of the achievement of that learning.

Cognizance
Fifth, knowledge-based teaching, so that those who assist students in learning are knowledgeable themselves. Right now there is a myth—maybe not a myth—that all you have to do to be a good a teacher is to “love to teach,” which is about as smart as saying that all you have to do to be a good surgeon is to “love to cut.”

Democratic Dynamics
Sixth, democratic dynamics. Without democratic dynamics evident around them, children will not learn how to practice democracy. You can preach about democracy until you’re blue in the face, you can have them read all the chapters and all the books about democracy, but if children don’t see adults in their lives interacting in democratic ways, they will not believe democracy is important or possible for all.

Safety and Discipline
Seventh, safety and discipline. It’s tough to concentrate on the substance of learning when you fear that the kid next to you might pull a gun out or will stab you in school.

Incentives
Eighth, incentives. I don’t know why we assume that in schools, moral arguments are enough. Ray Marshall, former secretary of labor under Jimmy Carter, said that the problem with moral arguments is that they only work with moral people. We must inject into education incentives for people to do right by kids and disincentives for failure to do so. But they need not all be individual and they need not all be monetary.

And most importantly, we must also inject into education incentives for the learners as well, so kids don’t come to the classroom and sit down with the attitude, “I’m here, now you learn me.” This is the only country in which I have lived where the assumption is that success is primarily the result of innate ability and circumstances. Everywhere else the assumption is more that it is primarily the result of effort. The worst thing that we can do is teach children that their fate is not within their hands; and the best thing we can do, especially for children who are vulnerable in our society, who already have odds against them, is to arm them with the conviction and confidence that they can change their fate and their circumstances by trying hard, by making an effort.

Resources
Ninth, necessary resources and autonomy. At a time when the needs of children are increasing, the resources to assist them are decreasing. To boot, educators have little confidence that the control freaks at the headquarters will allow them to implement what they labor to develop.

Accountability
And finally, shared accountability. Let’s not forget that schools cannot be an oasis of accountability in a desert of apathy and indifference. Schools can’t do it alone. We can’t hold educators alone accountable for children’s learning. Children spend approximately 9 percent of their year in school, the other 91 percent outside of school. So hold teachers responsible for that which is in their control, to know their stuff, to teach it well to all children, and to participate in nurturing the children’s readiness to learn. But not for that which can only be within the responsibility of parents, nor that which can only be within the responsibility of health and social service agencies. Shared accountability is the context for success for standards, for assessment, for learning, for children.

Let me also share with you my sense that moving the standards and assessment agenda will be no easier than moving any other aspect of the reform agenda. This is very tough. It means that schools will have to change. Yet, the general public holds suspect any school that doesn’t resemble the school that they remember. They say, “There is nothing wrong with these schools, I went to these kinds of schools that you say must change and look how well I turned out. What these little buggers need is a swift kick in the you-know-where.”

We [educators] liked schools so much that we decided to hang around them for the rest of our lives.

And we, the educators, are even worse about it because we are precisely the ones who were most successful in the kinds of schools that we must now change. We liked schools so much that we decided to hang around them for the rest of our lives. So it’s a tough, uphill battle and even now, after a decade of reform efforts, my colleagues in my district—in my union—say to me, “I’ll support reform after we stop violence in schools, after we get class sizes down to where they ought to be, after we get the necessary supplies and the right support,” and so forth. And I say to them, of course, if we got all these things right, why would we need reform?

Selling the Idea of Reform
Reform, we try to convince our colleagues, is something you do not after
you get things right, but something you do in order to get things right. And it’s a tough sell. Because while you’re here at this conference discussing assessments and standards, teachers are worried about what they will do tomorrow morning just to get through the day. I would urge you to listen to those voices, and to listen to other voices, those of your colleagues and mine, who have meaningful questions to raise about this reform effort.

Will these assessments tell us anything about whether or not children are also learning “habits of the mind?”

Many of these educators are people I admire a great deal. Linda Darling-Hammond for example, would caution us that taking the temperature of a patient will do nothing to cure him, no matter how often and how accurately you take that temperature. She would argue that assessments ought to serve as information—not immediate triggers for consequences. Debbie Meier would ask: Will these assessments tell us anything about whether or not children are also learning “habits of the mind?” Will these assessments measure kids’ ability to think? Will they tell us if they’ve learned how to think, how to construct their own meaning, how to raise essential questions, how to see connections, how to discover patterns, how to solve problems, how to make good decisions, and how to work cooperatively with others? Or will they be simply de-contextualized content in a strict sense of knowing formulas and simple answers to simplistic questions?

UCLA’s own Jeannie Oakes might raise other very important, intelligent questions. She would ask: What about ethical considerations, what about moral decisions, and what about the question of whether or not children have learned how to make moral choices at a time when increasingly we know more than we ought to choose to do?

Quality

And Grant Wiggins would ask: Is it a question of developing the right tests or is it a question of developing the right systems and processes and dynamics—not just events or acts or instruments? And he uses the football analogy: Do we know that the Buffalo Bills are a great football team just because they beat Dallas last Sunday? Shouldn’t we wait to see how they do throughout the entire season?

So, shouldn’t we think of assessments of student learning and progress toward standards as more complicated than we do now? And shouldn’t the real question be quality, i.e., whether or not students are doing quality work and achieving quality levels? I would add my own cautions: that we try to resist the temptation to unnecessarily polarize the issue of standards and assessments, to unnecessarily feel compelled to choose between standards and standardization, between local and national, between old and new, between depth and breadth, between rigor and rigidity, between teaching and testing, between equity and excellence.

...the challenge is to make education successful for those for whom it has never been successful...

Goals of Reform

I would say that the goal of reform is the same as the goal of standards and assessment, and that is not to make more mere exceptions but rather to make the exceptions into the norm. So if we cannot have standards that are achievable by all students, or at least most students, then I would raise the question of whether or not they’re worth having. Because excellence without equity is not excellence. Equity is a necessary component of excellence. And the challenge is not to find additional ways to help those whom God has already helped; the challenge is to make education successful also for those for whom it has never been successful, and for whom schools have never been hospitable places for learning.

Later on this month, we will hear about a major poll done by Lou Harris. I served on the advisory team for that poll, which raised this question to teachers. Do you support standards, high and rigorous standards, if they do not include some assurance that they are achievable by most kids, particularly the most vulnerable kids? Overwhelmingly, teachers said no, we do not. It is clear that teachers believe that these two issues are connected.

Opportunity to Learn

So I would say, do put a premium on both at the same time: high and rigorous standards and increasing opportunities for all students to meet those standards. Let us not wait for one to occur before the other begins; they can happen in tandem. Secondly, do put a special emphasis, as many have at this conference, on teacher development, because the Holmes Group report was right to suggest that teachers teach as they were taught. And if teachers do not themselves know the importance of standards or have the means to help students achieve high standards, students will not achieve these standards.

And lastly, I would caution against placing high stakes on even authentic assessments too early on. Probably the surest way to kill this worthwhile effort is to place high stakes early on. Real innovation, like real learning, does not happen under threat or in an atmosphere of fear. To do this right, we must do what we should do for all reform and that is, ensure an environment safe for innovation. In other words, we must admit that reform is a search, that it’s an inductive process—that along the way
there will be false starts, wrong turns and negative findings; and that negative findings are no less valuable to us than positive findings. AFT President Albert Shanker has a good way of expressing it. He said, when we try something and it doesn’t work we have a responsibility to publicize it “so that nobody else dies of the same cure again.”

So let’s find standards and authentic assessments—the kind that cater to good pedagogy and real learning, not just to convenience, not just for the purpose of tracking and sorting and labeling students permanently and early on. What is good assessment and what are good standards is the same question as what’s good curriculum and what’s good education. Good standards and good assessments ought to be a mirror of the curriculum and education that children have.

Churchill said, “Success is nothing more than going from failure to failure with undiminished enthusiasm.”

This will be tough because real change is real hard and takes real time. But even if we fail along the way, we should take Winston Churchill’s advice about the definition of success. He said, “Success is nothing more than going from failure to failure with undiminished enthusiasm.” So although it may be too soon for an emerging consensus on how to even frame the questions on standards and assessments, conversations such as those I have witnessed at this conference can and will help. And I for one have detected enough common ground to give me hope. I hope you did too.

The following new reports will provide initial understandings and valuable lessons for those interested in performance assessment at the classroom and school level. Contact Kim Hurst at 310-206-1532 to order these or any other CRESST reports, videotapes, or other products.

Writing What You Read: A Guidebook for the Assessment of Children’s Narratives
Shelby Wolf and Maryl Gearhart
CSE Resource Paper No. 10, ($4.00)

Designed to help teachers think about the important role of assessment in guiding students’ narrative writing, this guidebook should enable teachers to draw close connections between curriculum, instruction, and assessment. Derived from research that focused on classroom practice and the development of assessments informative to teachers and students, the guidebook focuses on the assessment of narrative. Educators should teach narrative, suggest the authors, not as an ever-shifting set of lovely stories to be lauded, but as a foundation for analysis, reflection, and criticism which can, in turn, be used as a resource for children’s original writing.

The guidebook provides specific examples for the scoring of student writing, including a narrative feedback form and narrative rubric to help teachers evaluate students’ understandings and learning development. Narrative areas of theme, character, setting, plot, and communication, as well as two concepts generic to all writing, convention and writing process, are illustrated and explained. The authors provide specific illustrations of student writing and teacher-developed assessments.

Whose work is it? A Question for the Validity of Large-Scale Portfolio Assessment
Maryl Gearhart, Joan Herman, Eva Baker, and Andrea Whittaker
CSE Technical Report 363, ($3.00)

Criticisms of standardized assessments of children’s writing have focused on issues such as the limited time provided to students to write, the artificiality of writing topics and assignments, and the restricted genres of writing assessed. Responses to these criticisms have prompted a move toward further authenticity—performance-based assessments that may incorporate shared readings of common background texts, collaborative planning, and opportunities for students to revise their work.

Based on an in-depth analysis of nine elementary teachers actively using writing portfolios in their classrooms, the researchers of this study focused on the question of portfolio authorship, that is, “Whose work is it?” If students receive assistance from parents, teachers, and classmates, then when raters assess student portfolios, whose work are they assessing?

The authors conclude that if portfolios are used to rank or make serious decisions about students, schools or districts, portfolio ratings should be adjusted to reflect differences in levels of support and assignment difficulty.

(continued on page 7)
On October 5, 1993, CRESST and the UCLA Graduate School of Education hosted two distinguished Australian visitors, Virginia Chadwick, minister for school education and youth affairs, and Ken Boston, director-general of school education, New South Wales, Australia and chair of the Australian National Standards Panel.

Attending a reception at the UCLA faculty center and meeting with the CRESST staff, Minister Chadwick discussed the many similarities between school reform efforts in the United States and Australia. Australia has entered the fourth year of a major school reform program, and there is great interest in creating a broad framework for a national curriculum. “The discrepancy of two peoples separated by a common language does not exist when our discussions center on education reform,” said Chadwick.

During his visit, Boston met with CRESST and members of the Quality Education Design Consortium of the Los Angeles Unified School District.

Boston and Chadwick are especially interested in finding out more about the Goals 2000 program and school-to-work transitions in the United States. Programs developed to address issues of student behavior and school violence were stated as areas of special interest by Boston.

“It is only too often that we operate in our own small world,” said Ted Mitchell, dean, UCLA Graduate School of Education. “We are thrilled that we have so many ideas in common with Australia and that our discussions will lead to increased communication and educational improvements in both countries.”

Following their UCLA meeting, Minister Chadwick and Dr. Boston hosted Art Express in San Francisco, an exhibit of 200 of the best projects submitted as part of the Australian Higher School Certificate. Bob and Joyce Linn, Eva Baker, Harry O’Neil, and George and Anne Madaus represented the United States measurement community at the exhibit and were joined by school administrators and public policy officials from the Northern California area.

---

**CSE/CRESST Reports**

*...from page 6*

**Dilemmas and Issues for Teachers Developing Performance Assessments in Mathematics**

*Robert J. Flexer and Eileen A. Gerstner*

CSE Technical Report 364, ($4.00)

This report examines some of the dilemmas and issues that arose during the first two terms of work with teachers participating in a mathematics performance assessment development program. Additionally, the authors report on changes in teachers’ instruction and assessment as a result of the project.

During the study, many dilemmas and issues arose that were unique to each of the three schools studied, but the most challenging problems across all schools were the teachers’ focus on what was important to teach, and therefore assess; and how children could learn what was taught—all within the constraints of limited teacher time. As expected, preliminary results of the project were mixed, but hopeful. Researchers believe that future development and implementation of performance assessments in these classrooms hinge on teachers’ beliefs in these assessments as useful and practical tools.
Memorials

Desmond Nuttall

It is with great sadness that we report the passing of Desmond Nuttall, professor and chair of curriculum and assessment studies at the University of London’s Institute of Education.

Originally receiving his Ph.D. from Cambridge University, Desmond became a well-known international expert in educational assessment. He was a Fellow of the British Psychological Society, consultant to the OECD Project on International Education Indicators, associate commissioner of the National Commission on Education, and director of the Monitoring Procedures Based on Performance Variables Project, and co-director of the Evaluation of Youth Development Projects. He wrote widely on topics in assessment and school evaluation.

A close personal friend and associate to many of the CRESST research staff, Desmond visited CRESST several times over the past few years, most recently presenting at the 1992 CRESST assessment conference. His knowledge, expertise, and personal enthusiasm for improving educational assessment will be greatly missed.

Contributions may be made in Desmond’s name to Greenpeace (USA), 1436 U Street NW, Washington, DC 20009, or Greenpeace (UK), Canonbury Villas, London, England, N1 2PN.

Anneli Vahapaassi

It is our sad duty to report Anneli Vahapaassi died this year of cancer. Anneli’s speciality was language development in Finland, and she was the author of one of the major text series used in Finnish schools. Our association with her began with the IEA Study of Written Composition in 1980. During the next seven years, with her leadership on the International Steering Committee, IEA members confronted a number of policy and technical issues of continuing relevance to international comparisons of performance. Although her scholarship was a strong focus, Anneli was an especially good friend and colleague, an enthusiastic, warm, and optimistic person. Over the years she visited the United States, and UCLA, repeatedly. We will all truly miss her.

Contributions may be made in Anneli Vahapaassi’s name to Greenpeace (USA), 1436 U Street NW, Washington, DC 20009, or Greenpeace (UK), Canonbury Villas, London, England, N1 2PN.

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

UCLA
CRESST
Graduate School of Education
405 Hilgard Ave.
Los Angeles, California 90024-1522

ADDRESS CORRECTION REQUESTED  EF-13