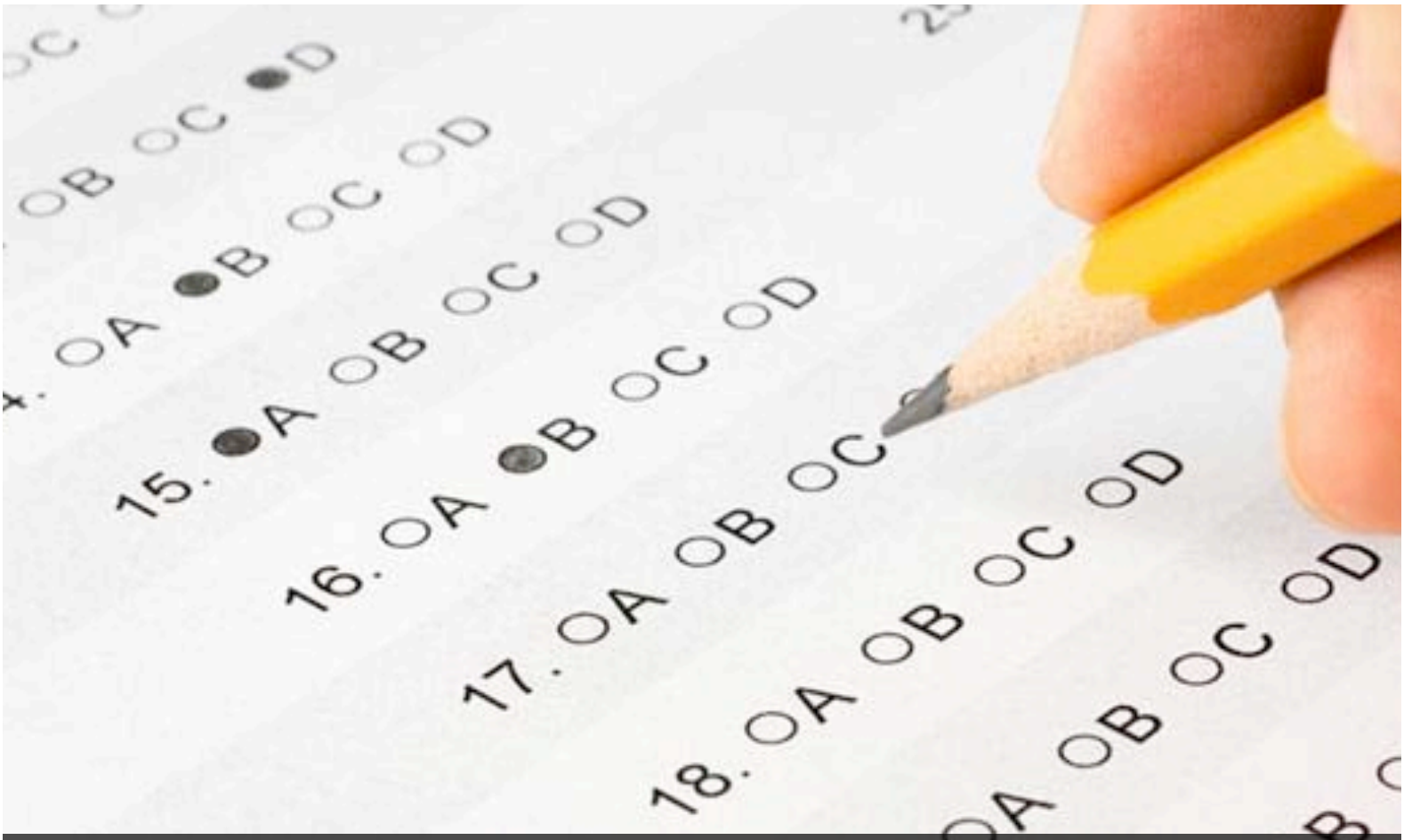


Working Toward an Assessment System with Value: Informing and Strengthening STEM Teaching and Learning

Summary of a California Teacher Advisory Council (Cal TAC) Symposium

March 2011

Working Toward an Assessment System with Value: Informing and Strengthening STEM Teaching and Learning

Symposium Summary

Held:
October 21, 2010
Sacramento, California

Convened by:
California Teacher Advisory Council (Cal TAC)

Co-sponsored by:
California Council on Science and Technology (CCST)
Center for the Future of Teaching and Learning (CFTL)

With generous funding support from:
The S. D. Bechtel, Jr. Foundation

Acknowledgements

CCST would like to thank the representatives from the California Teacher Advisory Council, Center for the Future of Teaching and Learning, and the S.D. Bechtel, Jr. Foundation who provided data and review throughout this process. We would also like to thank Nicole Lezin of Cole Communications for writing this report.

Copyright

Working Toward an Assessment System with Value:
Informing and Strengthening STEM Teaching and Learning
California Council on Science and Technology
March 2011
ISBN 978-1-930117-43-3

CCST is a non-profit organization established in 1988 at the request of the California state government and sponsored by the major public and private postsecondary institutions of California and affiliate federal laboratories, in conjunction with leading private-sector firms. CCST's mission is to improve science and technology policy and application in California by proposing programs, conducting analyses, and recommending public policies and initiatives that will maintain California's technological leadership and a vigorous economy.

The California Teacher Advisory Council is a group of highly accomplished K-14 teachers with demonstrated expertise in science, technology, engineering and/or mathematics (STEM). Cal TAC was created by the California Council on Science and Technology (CCST) to provide perspective and 'wisdom of practice' to policymakers and others. Cal TAC members work together to strengthen student learning in the STEM disciplines. Specifically, Cal TAC provides a much-needed connection between the classroom, education experts and policymakers with regard to high quality teaching within the STEM disciplines.

For questions or comments on this publication contact:

California Council on Science and Technology
1130 K Street, Suite 280
Sacramento, CA 95814
(916) 492-0996
ccst@ccst.us

Table of Contents

MESSAGE FROM CCST AND CAL TAC

The California Teacher Advisory Council (Cal TAC) was created in 2005 as a means for bringing real-world classroom experience — the "wisdom of practice" — to policymakers and others whose decisions affect the quality of science and math education in California. As part of their work, they hold annual symposia on issues critically important to effectively teach math and science to California's students. In October 2010, Cal TAC held a one-day symposium to discuss the implications of formative and summative math and science assessments for teaching and learning in California classrooms. Invited guests included science and mathematics teachers, school administrators, researchers, district and state agency education officials, legislative staff, parents, and advocates.

By definition, formative assessments are those where students and teachers use immediate evidence of learning to adapt educational practices to meet learning needs minute-to-minute and day-to-day. In contrast, summative assessments occur at a specified period of time and have three purposes: to rank, certify, and provide grade level information and placement. Put simply, formative assessment is characterized as assessment *for* learning while summative assessment is assessment *of* learning. Both types of assessments are important for improving education but for different purposes. The goal of the symposium was to highlight innovative practices of both types of assessments, clarify the purposes and appropriate uses of each, and engage in a discussion of how quality assessments can be used to inform teacher professional development. We hope you find this summary report useful in your work and look forward to continuing the discussion of how best to strengthen student assessments in math and science. Key findings and recommendations are included below.

Key Findings

- California needs a coherent *system* of assessments because no single assessment can measure everything that is important for evaluating and improving education.
- Formative assessments are as, if not more, important for improving instructional practice as summative assessments, yet are currently not given as much attention, support, and time.
- Teachers have little formal training (in their initial preparation or in-service professional development) on how to effectively use formative assessments in their classroom.
- Summative assessments used for accountability purposes have moved beyond their intended use—the heavy emphasis on raising summative assessment results has negatively impacted teachers' ability to teach math and science.

Recommendations

In light of these findings, we strongly encourage education leaders and policymakers to take the following actions:

State-wide Education Leadership and Policy Organizations

- Move towards an innovative computer-based system of summative assessments which can provide for more valuable assessment tasks that can better reflect students' skill level, including problem-solving and the application of scientific/mathematic concepts, and provide timely results to teachers so that they can adapt their instruction appropriately. We are fortunate that states have paved the way in this initiative and can look to Oregon and Minnesota as models.
- Include current classroom teachers in decision-making process related to summative assessment strategies and practices.

Teacher Preparation Programs

- Reflect the importance of effective formative assessments and their utility in teaching practice by embedding effective practices in teacher preparation programs curricula.

School District and Site Leaders

- Encourage collaborative professional development on formative assessment practices at school sites.
- Accept responsibility with teachers, principals, and parent groups like the PTA, in educating school communities on purposes and uses of different assessments.
- Include teachers in decision-making process related to both formative and summative assessment strategies and practices.

Philanthropy

- Support efforts to create a web-based clearinghouse of effective formative assessment practices that can benefit teachers statewide. Access to the clearinghouse should be free and include the capacity for teachers to share stories and evidence of their experiences with the practices.

Introduction

Picture a classroom in which students are quietly filling in the ovals on a standardized test, some chewing on their pencils as they concentrate and struggle to find an answer, some staring out the window, and some racing from one question to the next with confidence.

In another classroom, students are taking a different type of test — on a computer, with animation that simulates different aspects of a word problem. Depending on each student's responses, the test automatically adjusts its level of difficulty.

Elsewhere, students might be grasping electronic clickers in their hands with the enthusiasm normally reserved for video games. They gaze up at a screen to see how their answers compare to those of their classmates, while the teacher reacts to an instant and revealing snapshot of whether the students are ready to move on.

Later that day, the teacher carefully watches the interaction of students working in groups to solve a problem, noting the types of questions they ask and the obstacles they encounter and overcome.

All of these are examples of different ways that teachers assess student learning — and respond to what the various assessments reveal. Even though these are constant, routine events in classrooms across California, they are rarely considered as part of a coherent system that connects teaching and learning.

How do we assess the learning that occurs — or doesn't occur — in California classrooms? How can we do a better job of assessing student learning more accurately and usefully, so that teachers and school systems can be more responsive to evolving student needs? What form would better assessments take, and what can we learn from other states that have tackled the same issues?

The California Teacher Advisory Council (Cal TAC)

These questions were at the heart of a symposium convened by the California Teacher Advisory Council (Cal TAC) in October 2010. Cal TAC was formed in 2005 by two co-sponsoring organizations: the California Council on Science and Technology (CCST) and the Center for the Future of Teaching and Learning (CFTL). CCST and CFTL joined forces to form Cal TAC in 2005 as a means for bringing real-world classroom experience — the "wisdom of practice" — to

policymakers and others whose decisions affect the quality of science and math education in California.

CCST is a nonpartisan, impartial, not-for-profit corporation established over 20 years ago to offer expert advice to the state government and to recommend solutions to science- and technology-related policy issues. CCST's core support comes from California's major post-secondary institutions, which provide important backing, support, and resources to CCST. CCST is governed by a Board of Directors composed of representatives from its sponsoring academic institutions, from the corporate and business community, as well as from the philanthropic community. Together, these members are helping both the public and private sectors find answers to the important science and technology-related issues facing California.

CFTL is a public, not-for-profit organization dedicated to strengthening teacher development policy and practice. With its research partners, the organization promotes effective teacher development through data collection and analysis, state and national policy, and legislative initiatives.

About the Assessment Symposium

The day-long symposium was an opportunity for science and mathematics teachers, school administrators, researchers, district and state agency education officials, legislative staff, parents, and advocates to discuss the implications of different types of assessments for teaching and learning in California.

To open the symposium and frame the subsequent discussions, Joan Herman, Director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA, delivered the **keynote address**. Dr. Herman's presentation was followed by two panel discussions: one on innovative **formative assessments**, and a parallel one on **summative assessments**.

Each of these components of the symposium is summarized below; an agenda and list of participants are provided in Appendices A and B.

Comments from a "Give Your 2 Cents" exercise appear like this throughout the summary.

In addition to question-and-answer discussions that followed each presentation, symposium participants also were encouraged to contribute their **"2 cents"** worth of ideas and questions in writing. Although there wasn't enough time during the symposium to discuss these, they are presented throughout this report with a 2-cent logo, as shown here.

The Time is Now: Assessment Systems Coherent with Learning

Keynote Address by Joan Herman, Director of the National Center for Research on Evaluation, Standards, and Student Testing

Joan Herman opened her keynote address with the observation that this is an exciting, propitious time for the educational assessment field, with major federal Race to the Top investments (to the tune of $170 million) in multi-state assessment consortia and the development of common core state standards in reading and math.

The unprecedented levels of resources and attention to this issue serve as incentives to get it right this time, Herman said. However, getting it right — or at least "righter"— depends on understanding what's wrong with the current system.

What's Wrong with Current Assessments?

Herman noted that the purpose of assessments is to communicate what is important — in terms of both teaching and learning — and to hold schools and teachers accountable for whether or not that learning takes place. Ideally, assessments yield the types of data that support improvement, giving students, parents, policymakers, and the community as a whole the information they need and can put to use.

Unfortunately, she said, current assessments are not measuring the right things. "Because we're not measuring the right stuff," she pointed out, "we hold teachers accountable for the wrong stuff." For the most part, she said, the data come back in formats that are not useful for the purpose of improvement — "too little and too late to inform ongoing teaching and learning."

Grain size: It matters at what curricular grain size we hold students and teachers accountable. When the grain size is small, instruction will drift toward the bits and pieces.

Getting it Right(er)

Instead of thinking of a single assessment that tries (and fails) to serve multiple purposes, Herman suggests a system of assessments, coherent with meaningful teaching and learning. As she showed in one of her slides (Figure 1), the federal Race to the Top Requests for Proposals (RFPs) recognize that assessments in the service of learning cannot be served by a single agenda. Although this recognition is a plus, the downside is that the expectations are still overly ambitious, she said.

Figure 1: **Federal Expectations**

Purpose	Assessment Type	Primary Users	Use
Accountability/ Evaluation	On-demand annual	State, District, Schools, Teachers, Parents, Students, Public	▪ School/Teacher/Principal effectiveness and capacity building ▪ Status/Growth re: college career readiness ▪ General Feedback curriculum strengths and weaknesses, student strengths and weaknesses? ▪ Recognize and build on excellence
Accountability/ Evaluation	Through Course, Curriculum Embedded	Same as above	▪ Same as above
Monitoring/ Supervision	Interim Assessment, Unit Exams, Curriculum embedded assignments	District /School Administrators, Teachers, Parents, Students	▪ Identify struggling students, teachers schools ▪ Diagnose/adjust? ▪ Identifying promising practices
Formative	Curriculum-embedded, student work, class discourse, discussion	Teachers, Students	▪ Inform immediate teaching and learning

Using images of a partial and full portrait to make her point, Herman contrasted the yield from an annual, relatively brief assessment (the incomplete picture) with the much richer, more comprehensive (and more accurate) picture that emerges from multiple kinds of assessments, conducted multiple times.

Today's tests, very limited, superficial in treatment of domains assessment

VS.

Tests that can capture the full domain, balanced and deep coverage

Four Kinds of Coherence

We need a *system* of assessments, Herman said, because no single assessment can measure everything that is important and useful for different users of the information. For example, policymakers might legitimately want a quick barometer of how students are performing (such as that offered by a single multiple-choice test), but those results don't help teachers adapt a curriculum and tailor what they say and do in the classroom the next day.

The system, moreover, needs to be *coherent*, moving everyone in the same direction (as long as it's the right direction, Herman added). Herman reviewed four types of coherence to keep in mind as assessment systems are revamped:

- **Fundamental coherence** — as a valid measure of significant learning. Herman noted that assessments for learning begin with a learning goal and then construct assessment tasks that elicit evidence of learning, using an interpretive scheme that enables us to view students' responses in ways that tell us where they are relative to the learning goal.

  Although this seems like an obvious sequence, Herman said, most large-scale tests do not begin with the learning goal. Instead, they start with the assessment tasks. Valid assessments, must not only be aligned with significant learning goals, but also accessible and unbiased, allowing all children to demonstrate what they know (and not, for example, confounding a lack of science knowledge with language proficiency problems).

- **Horizontal coherence** — the degree to which assessments inform improvements in learning. In order for this to be the case, assessments must have *diagnostic value*, providing information to both teacher and student alerting them to problems and giving some indication as to why there is a problem (so that it can be addressed). They also must be *sensitive to instruction* — i.e., measuring what students actually learn in school, rather than their basic intelligence (what they bring to school). "We want to know what school adds," Herman said.

- **Developmental coherence** — how assessments start with student learning goals and progressively feed into time intervals (daily, weekly, by unit, quarterly, annually, over the entire K-12 span). Ideally, Herman said, one could start far out along this chain, such as knowing where a student should be at the end of the year, and work back to determine a learning progression that starts with what he or she needs to be teaching today, and how this feeds into the next day, week, and so on. This approach allows a focus on the "big ideas" of any particular discipline and how they are developed across time, assessing the full range of what we want students to know and do along the way.

- **Vertical coherence** — putting all these pieces together, combining valid assessments with appropriate use of their results. Starting with the learning goal, vertical coherence calls for developing assessment tasks and measures of the "right stuff," interpretive schemes that help us understand students' status relative to the learning goal, and put that information to use by taking actions that further students' learning. One implication of quality assessments and how they are used is that they then require that teachers have the knowledge and skills to use these assessments in ways that improve students' learning. Many studies of formative assessments, Dr. Herman noted, have found that teachers often lack this knowledge.

Starting with the learning goal, vertical coherence calls for developing assessment tasks and measures of the "right stuff," interpretive schemes that help us understand students' status relative to the learning goal, and put that information to use by taking actions that further students' learning.

Reflections on Assessments and Their Use

Turning to the big picture, Herman said, the vision is of everyone — teachers, schools, districts, state departments of education — using a consistent set of information about student learning to address the same types of learning, and using this information to take concerted action that actually improves learning.

System validity. Like good assessments, Herman explained, good systems must start with a definition of purpose; those that try to serve all will end up serving none well. At this juncture, on the cusp of developing a new set of assessments, the question becomes whether we can move ahead in a better way.

To assess a system's validity, she proposed four essential questions to use as a checklist:

- Does it serve priority purposes?
- Do individual and collective measures provide accurate information for intended decision-making?
- Do these measures enable all students to show what they know and to demonstrate progress?
- Are the system consequences as intended?

System support for change. From the federal perspective, Herman observed, the purpose of these assessments is to ensure that all students are ready for post secondary education and generally prepared for success in life in an increasingly competitive global economy, which in turn suggests dramatically different expectations for teaching and learning. Ideally, assessments would help teachers understand and attain these goals.

However, Herman noted, "We all know that no matter how wonderful an assessment is, it's a small part of what has to happen to improve learning — a necessary, but not sufficient part." Teachers and teaching practice also have to move from where they are to where they need to be, she said, noting that the new assessments cannot simply be "dropped" onto teachers. "They will need systems for support and implementation."

"We all know that no matter how wonderful an assessment is, it's a small part of what has to happen to improve learning — a necessary, but not sufficient part."

Assessing the assessments. The coherence of assessments with significant learning begins with rich understanding and clear goals — and a good sense of how they develop, so that the assessment system will fit into a developmental trajectory.

Coherence also starts with priority purposes, Herman said, "putting money down on some places and walking away from others." It's a complex design problem with many competing tensions, she noted, calling it an "inquiry project" with no single right answer.

Finally, she said, validity requires evidence that the system and its elements are working — that they are serving their intended purposes, aligned, fair and accessible, reliable, instructionally useful, instructionally sensitive, developmental, comprehensive, educative, and model good practice. "We are all consumers, and we ought to be smarter and demand the evidence. We get a very meager set of evidence supporting our current assessments . . . We don't know as much as we need to about what is being assessed and how well we're assessing it."

We're it! Looking back over decades in the field and many waves of change, Herman lamented, "Each time, we thought testing was going to help fix the world. Each time, it hasn't done that — and has in fact caused negative consequences." She urged her audience to keep this in mind and closely monitor the negative consequences that may result. "We have to keep an eye on what's happening…We're it! The public, advocates, the people here — we need to keep our eye on the ball."

Formative assessment: Is it an equity issue?

A Focus on Innovative Formative Assessments

Assessment-Centered Teaching: A Reflective Practice

Kathy DiRanna, Center for the Assessment and Evaluation of Student Learning, WestEd K-12 Alliance

Kathy DiRanna reviewed the framework behind Assessment-Centered Teaching (ACT), which has been developed through a National Science Foundation-funded collaborative, the Center for the Assessment and Evaluation of Student Learning (CAESL). WestEd, UC Berkeley's School of Education and Lawrence Hall of Science, CRESST at UCLA, and Stanford University are all members of the CAESL collaborative.

Assumptions about Teaching and Assessment

CAESL's work involves building a community of practice that more closely links knowledge (in the form of assessments) to those who use the information (teachers). In developing the ACT framework, DiRanna explained, she and her colleagues at CAESL realized that they were making numerous assumptions about the instructional and assessment practices of teachers, but many of these turned out to be erroneous. (And not through any fault of the teachers, she hastened to add.)

The assumptions included the idea that teachers who have good *instructional* practices naturally also have good *assessment* practices (which is not necessarily the case, if teachers were not trained to do so and don't have the time in their class schedules to have the opportunities to conduct assessments).

Another assumption was that if teachers know what students' misconceptions are, they'll be able to respond to them. Again, this only works if teachers have been taught how to elicit accurate information about students' learning needs and have opportunities to do so.

Since the value of feedback is well-known, DiRanna said, they also assumed that teachers routinely provide it — again, not so. Moreover, even if they are able to analyze and interpret assessment results, teachers do not necessarily have the knowledge, skills, and capacity to provide the appropriate instructional "next steps."

The ACT Framework

The CAESL ACT framework that DiRanna described is designed to help teachers plan for assessment, think about what they're teaching, analyze students' work,

and then guide and shape assessments based on what they learn, in a continuous loop. It is centered around four tools:

- **Developing Conceptual Flow.** This gets at the big ideas and concepts — how they are nested within others, which supporting concepts they might be linked to, where they lead. These guidelines are important for teachers, but even more so for assessments.

CAESL developed a parents' assessment toolkit with online tools, workshops, handouts and policy briefs.

- **Creating a Record of Assessment in Instructional Materials (RAIM).** RAIM analyzes how instructional materials support assessments. For example, do questions in a book chapter really address the material for which teachers and others want students to be responsible? Are there better ways to use instructional materials? The result should be an assessment plan that includes a pre-assessment, juncture assessment, and post-assessment. With this closer look at what students are learning, teachers can then flag more specific assessment points that help them figure out where the conceptual pitfalls are for students within each lesson. That's why the term "juncture" is used: if students don't have an understanding of certain concepts at a particularly crucial juncture, then they won't be able to build on that knowledge and move on.
- **Converting Expected Student Responses (ESRs) to Rubrics.** DiRanna asked her audience to quickly list five characteristics of living things. The pop quiz yielded many different responses and illustrated how these could be turned into rubrics that capture more details about what the student already knows at different levels, what an expected response might be, and what the student needs to learn. If the teachers develop a more detailed understanding of the expected student responses, they can be more attuned to the signals the student's work is sending about their conceptual understanding of the material (or lack thereof).
- **Documenting a Complex Record of Assessments.** At the end of the CAESL process, teachers had developed a much more complex picture of what assessment systems could explore, with not only more elements but more relationships among them.

In closing, DiRanna noted that the current environment is hostile to quality formative assessments because they are time-consuming and the quality of materials is poor, gravitating to the lowest common denominator. Still, DiRanna remains hopeful, seeing great potential and possibility — although, as she put it, "We are not there yet . . . The more we get into this, the more we learn."

Next Generation Assessments in Mathematics

David Foster, Silicon Valley Mathematics Initiative

David Foster is Executive Director of the Silicon Valley Mathematics Initiative, a 41-member organization made up of school districts and charter school networks that was formed in 1996 to support the teaching and learning of mathematics.

> "**Formative assessment** is students and teachers using evidence of learning to adapt teaching and learning, to meet immediate learning needs minute-to-minute and day-to-day."
>
> *— Dylan William*

Defining Formative and Summative Assessments

Quoting from the Annenberg Institute of School Reform, Foster defined formative assessment as:

> "the process of studying student work in a meaningful and challenging way to be data-driven, to reflect critically on our instructional practices, and to identify the research we might study to help us think more deeply and carefully about he challenges our students provide us . . ."

Summative assessments (such as high school exit exams or SATs) serve three purposes, Foster noted: to rank, certify, and provide grade level information and placement. They are important, but distinct from formative assessments. (More frequent or interim summative assessments, such as chapter and unit tests, Foster pointed out, can be thought of as "mini-summative assessments.") "Those aren't really formative unless they drive instruction…Unless the teacher makes a different decision, it's not formative, by definition."

Foster noted that in any discipline — social studies, science, or math — researchers confirm the importance of using student work and formative assessment to guide instruction. Despite this body of evidence, though, teacher training programs do not provide in-depth courses on formative assessments.

To emphasize the ongoing and instruction-driving aspects of formative assessment, Foster quoted Dylan William from the University of London: "Formative assessment is students and teachers using evidence of learning to adapt teaching and learning, to meet immediate learning needs minute-to-minute and day-to-day."

Formative Assessment Strategies and Cycles

In "Inside the Black Box," an article he co-authored with Paul Black, William lists five effective formative assessment strategies:

- Clarifying learning intentions and sharing criteria for success (echoing the points made in earlier presentations about a clarity of purpose),
- Engineering effective classroom discussions (particularly difficult in mathematics classes, because of the traditional paradigm of how math classes should be),
- Providing feedback to move learners forward,
- Activating students as owners of their own learning, and
- Activating students as instructional resources for one another.

Recent research shows that with professional development categorical flexibility, money is being spent elsewhere. The positive consequence of this is that school sites are now relying upon the expertise of their own teaching staff and working together . . . promising for the spread of formative assessment.

The last two of these strategies, Foster noted, are centered squarely on the student. All five strategies guided the Silicon Valley Mathematics Initiative's work by stimulating them to think about a formative assessment cycle, with teachers posing questions to challenge students, listening to their responses, examining their thinking, and then — the most crucial part — changing what they do next.

In the traditional classroom, if a 10-point quiz revealed widely different levels of understanding and mastery of the material, one of three things would happen:

- The teacher would re-teach the topic (slower and louder than the first time, Foster predicted) for the entire class.
- The teacher might identify a few students who need extra help, and work with them while the rest of the class moves on (if he or she could somehow do this with 32 different children).
- The teacher, recognizing that time is of the essence, moves on to the next topic.

"Obviously," Foster noted, "none of those feels good."

Another Option: Re-Engagement

After a decade of working on this problem, Foster and his colleagues developed a re-engagement model. Instead of a binary mode typical of American standardized assessments (e.g., a student either knows the answer or doesn't), re-engagement attempts to reveal information that is useful to the teacher in terms of where to go next.

In this model, teachers create tools that help students look through the lens of other students' thinking, raising the level of cognitive demands made of them. For example, if students are presented with a math problem that involves proportional reasoning (see box) — an important idea as students move from elementary to middle school math — they can use each other's misconceptions to understand why that misconception occurred, and to explain how the concept and chain of reasoning actually works. The approach helps teachers make sense of why students are reasoning a certain way, and how to respond. ("Teachers aren't always good at it," Foster observed, "but with training, they can get much better.")

The Candy Problem

Anthony makes candies. First, he mixes 1 cup of cream with 2 cups of chocolate. In all, he uses 9 cups of these 2 ingredients. How many cups of chocolate does he use in this recipe?

In re-teaching, Foster summarized, teachers teach a unit again to address basic skills that are missing by doing the same or similar problems over and practicing them so that students learn the procedures. The focus is on underachievers, and the cognitive level is lower.

In re-engagement, on the other hand, the teacher (along with the students) revisits the student's thinking in order to address conceptual understanding, examining the task from different perspectives. Students' approaches and solutions are critiqued not from the perspective of whether they were right or wrong, but to make connections — for an individual student, and to engage the entire class. As a result, the cognitive demand is higher.

Comparing CST and MARS

Foster's group has analyzed years of testing trends, by grade level for mathematics classes in grades 2 through 8. In comparisons of test results, the analysis revealed a large number of false positive — i.e., in eighth grade geometry, nearly half of the students who appeared to meet grade-level standards on the California Standards Test (CST) (44.5%) could not pass the performance-based exam, when they had to find answers without relying on multiple-choice options. Yet not a single student who met performance standards according to the criteria of the Mathematics Assessment Resource Service (MARS) failed to do so on the CST.

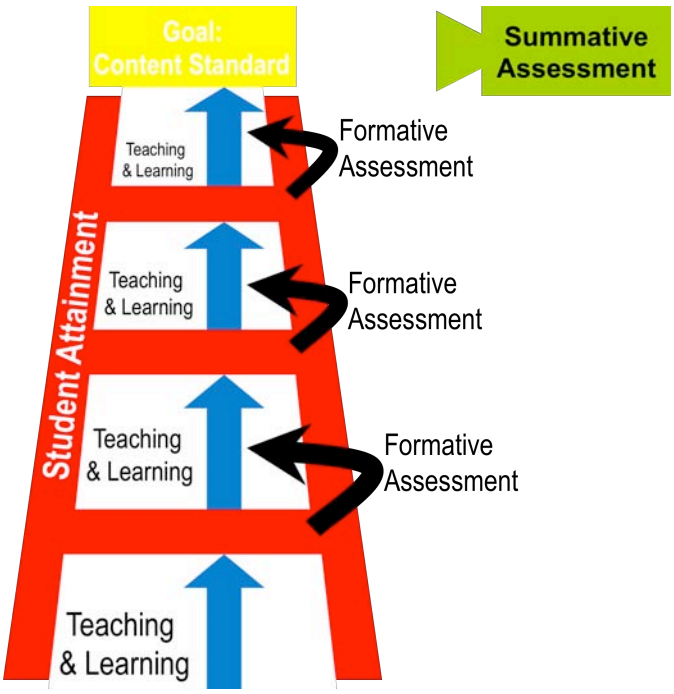
Resources on Re-engagement

In closing, Foster offered a number of resources on this topic, including video footage from classrooms throughout the Bay Area depicting many of the re-engagement scenarios he described. These are available through the Inside Mathematics Web site, www.insidemathematics.org.

A Teacher's View of Formative Assessment

Alastair Inman, Lexington Junior High School

When Alastair Inman learned that he would be speaking about the teacher's view of formative assessment at the symposium, he began asking his colleagues about that the term meant to them. The answers varied:



"It's anything that's not summative,"

"If it's not a final test, then it's formative,"

"It's anything you don't grade, if you put a grade on it, it's summative."

Inman likes the definition that David Foster used in his presentation moments earlier: "Activities that teachers and students undertake to get information that can alter teaching and learning."

To illustrate how different assessments fit together, Inman used the image of a ladder. At the top, he explained, are state standards — which represent the teaching goal of what we're trying to get students to learn. Teaching and learning in the classroom move students up the ladder to attain the goal. A summative assessment at the top tells us whether they have achieved the goal, as defined by content standards.

> "Assessing each individual student's level of attainment and differentiating individual instruction based on those assessments becomes increasingly challenging as class size increases."

All along the rungs of the ladder are formative assessments that assess each level of attainment and help both student and teacher determine where the students are, and what they might need to keep moving in the right direction. Feedback on student attainment — constant feedback — is what keeps the student on track.

Times have changed since Socrates could ask questions of his small group of students gathered under a tree; just in the past decade, Inman's classroom has grown from the high 20s to the 30s and even into the 40s. "As a teacher," he said, "when it comes to formative assessment, class size matters." Assessing

each individual student's level of attainment and differentiating individual instruction based on those assessments becomes increasingly challenging as class size increases.

Formative Assessment Techniques

> The teachers in this room today are obviously all effective, high-quality instructors. How do we build out this expertise to all or more teachers?

Ideally, Inman says, formative assessment begins before the teaching itself. Inman described how before every major unit begins in his science classroom, he starts with a simple question: "Whaddaya know?" Whaddaya know about plants? About cells? About evolution? Whatever the topic may be that day, students sit in small groups and come up with some ideas, which are shared with the larger group and recorded on a whiteboard. "It gives me feedback about where the students are," said Inman. "If I find out they're on a higher rung of the ladder," he continued, returning to his ladder analogy, "why teach as if they're on the lowest and waste a couple of days?"

The opposite could happen as well — instead of discovering that his students are already advanced, he might discover that they are starting with misconceptions that need to be addressed.

After the "Whaddaya Know?" assessment, a variety of other forms are used on an ongoing, continuous basis, providing feedback on every activity. This approach is challenging, because it requires finding time to re-teach as needed, without falling behind the steady pace required to get through the unit. It's understandable that teachers feel pressured to continue, Inman said, even if they feel some of their students are not ready.

Inman relies on many different techniques to assess his students' learning. If the same students are always raising their hands to answer his frequent questions, for example, he allows extra time for others to raise theirs. He may also call on those who have not raised their hands, or encourage sharing in pairs. Sometimes he relies on thumbs up or down, or short written answers (either on paper or whiteboards).

A popular technique is the use of electronic clickers. Each student gets one; they respond to multiple-choice questions and their answers are collated on a computer, showing up on a screen as an instant bar graph. "I get to see how the whole class is responding," Inman said, "and I get to see what mistakes they're making, so I can address misconceptions imbedded in their mistakes." In addition to instant feedback, he sometimes uses the clickers for more extensive written questions.

Clickers have another advantage, besides being fun for students and providing instant feedback for teachers: they reveal a need for help without embarrassing anyone. This is particularly important for junior high students, who may not want to ask for help in front of the entire class.

Inman uses technology in other ways, too — such as soliciting electronic "stickie" notes on a cyber wall that are read overnight and used in the next day's lesson. And he continues to find labs and experiments terrific opportunities for formative assessments because of the one-on-one interaction as he walks through the classroom. It also provides opportunities for mini-lessons if he sees signs of trouble, such as difficulty adjusting a microscope. "The 'whoa' I hear when they look under a microscope is one of the most satisfying things," Inman said. "It's immediate corrective action at its best!"

"Formative assessment is an integral, continuous part of teaching and learning that forms, shapes and guides everything I do in the classroom. Without it, I would be like the captain of a boat with my hand on the tiller, but going blind, with no idea which direction was forward."

Keeping a careful eye on his students as they answer questions or work in groups is a form of assessment in and of itself, Inman said. Subtle changes in facial expressions and body language signal that students are not getting it — as much as a carefully designed quiz might. Inman called this aspect of formative assessments "teacher radar" — being alert to a constant stream of information about "what kids do know, don't know, what they're feeling in the moment . . . a continuous adjustment in content, pace and tone."

In conclusion, Inman said, if an administrator entered his classroom tomorrow and suddenly banned summative assessments, he could live with that. "I'm not saying they're of no value," he explained, "but they're of value mostly to people who are not in the classroom all the time: principals and policymakers, rather than me and my students."

If the administrator banned formative assessments, on the other hand, "I wouldn't even know what that means or looks like," Inman said. "Because to me, formative assessment is an integral, continuous part of teaching and learning that forms, shapes and guides everything I do in the classroom. Without it, I would be like the captain of a boat with my hand on the tiller, but going blind, with no idea which direction was forward."

Reflections

Sandy Dean, Director of the National Board Resource Center at Stanford University, shared some of her reflections and reactions listening to the presentations on formative assessments. She was struck by the

assumptions Kathy DiRanna described about what teachers know, which echoed some of her own observations hearing from teachers who come to the Resource Center as part of the National Board Certification process. "A lot of research validates what teachers say," she observed, "but it's not in their hands." Many teachers don't even know the term "formative assessment." To Dean, this means we have to do a better job of helping teachers develop better practice. "It has implications for pre-service and for using research well," she said.

Professional development needs to include the "Why" (research), the "How" (strategies and techniques), the "Tools" (to enable the process and analyze the results), in a collaborative environment.

Several other questions resonated:

- What would happen if we started with assessment in mind?
- How can we incorporate the vertical path through teaching better in curriculum and assessment plans?
- How can we use the teacher's power — not in terms of being in control of what's happening in the classroom, but power in terms of using what's going on in the classroom to make decisions — given the current context that is so hostile to formative assessment?
- What is the student's role, and what are the practice and policy implications that lead to students not wanting to ask questions and demonstrate that they may not know something? ("The idea of listening to what kids say and do seems so simple," she said, "and yet, so often, it's not honored.")
- And last but not least, "Whaddaya know?" "Probably every conversation in education should start with that!"

A Focus on Innovative Summative Assessments

A Look into the Future: California's Summative Assessment System

Deb Sigman, California Department of Education

The Current Context

Deb Sigman opened her presentation by setting the context for California's current process for test development. She emphasized that California's summative tests are primarily designed to serve as accountability measures at the school and district levels — not as diagnostic measures within the classroom.

The cost of testing California's 4.5 million students[1] is high: approximately $76 million each year for development (through a contractor) and administration, and $25 million to districts and schools to offset the administrative costs of conducting the tests. In the wake of budget cuts, this has become even more challenging.

The contractors who develop California's tests are responsible for delivering tests that are valid (i.e., accurate in terms of their intended purpose), reliable, and fair (i.e., free from bias). Currently, the tests are paper-and-pencil, multiple choice tests (with essays added in grades 4, 7 and 10), largely for efficiency and cost reasons.

> The CST exams do not accurately reflect content. They assess testing ability.

The Common Core State Standards Initiative

In 2009, the Council of Chief State School Officers (CCSSO) and the National Governors Association Center for Best Practices develop standards for college and career readiness. This effort was voluntary and state-led, involving representatives (including parents, educators, researchers, and others) from 48 states, 2 territories, and the District of Columbia. The results were rigorous, research-based standards for English language arts and mathematics for grades K-12, tied to international benchmarks so that students will be competitive in a global economy.

Meanwhile, in California, Senate Bill 1 from the Fifth Extraordinary Session (SB X5 1) established an Academic Content Standards Commission (ACSC) to develop standards in English language arts and mathematics. Overall, these were to be largely consistent with the common core state standards (up to 85%), with the remaining 15% consisting of additional material. The Commission directed the

[1] This number represents only a portion of the total 6.2 million students in the state, since not all students are tested every year.

State Board of Education to either adopt or reject the recommended standards, in an effort to avoid tinkering.

Under the Sacramento County Office of Education, the ACSC was convened three times during the summer of 2010 to review the core standards for rigor and alignment to California's standards. Elements of the California standards were inserted, and the Commission then recommended that the State Board of Education adopt the common core state standards as amended, which they did on August 2, 2010 (the same day, not coincidentally, that the Race to the Top application required minutes from a State Board of Education meeting adopting the common core standards).

Assessment will be key to the new K-7 common core standards that prepare students for Algebra 1 in eighth grade, Sigman said, as well as for success in higher math classes. CDE is working on implementation plans for these standards, including frameworks, instructional materials, and professional development (as well as assessment), with the expectation that a plan would be presented to the State Board in November 2010. Sigman anticipates that significant changes will be needed to the Elementary and Secondary Education Act (ESEA) assessment requirements as a result.

Partnership for the Assessment of Readiness for College and Careers (PARCC) Consortium

Race to the Top included competitive RFPs for assessment consortia; two emerged from this process (Smarter Balanced, with Washington State as the fiscal agent, and PARCC, with Florida as its fiscal agent).

Transfer: The reason we need assessments that vary in terms of their curricular similarity to instruction is that we need to know, "How far will instruction travel?"

PARCC is a consortium of 26 states. California is one of the participating (as opposed to governing) states in the PARCC consortium. In September, PARCC was awarded $169.9 million, with an additional $15.9 million designated to help participating states (such as California) transition to common core standards and assessments, which would occur annually in grades 3-8 and once during grades 10-12.

PARCC's theory of action, Sigman explained, is to develop more meaningful standards anchored in the common core state standards (and thus consistent and comparable across states); higher-quality tests that measure critical thinking, problem-solving and writing in more sophisticated ways; through-course testing (in which students are tested closer in time to when they learned the materials, not just at the end of the year); and tests that take advantage of technology to provide more timely and varied options.

Ideally, these changes will lead to achievement results that are based on clear definitions of college and career readiness (so that students will have useful information in time to adjust), comparable results across states that are measured against consistent standards, accountability-driven improvements, and higher quality instruction (by giving teachers useful, meaningful, and timely information that they can use to adjust instruction and fine-tune lessons throughout the school year).

> Transfer and accountability seem at odds with one another. Accountability says link testing and teaching tightly. Transfer encourages us to move as far from instruction as we can.

The anticipated timeline, Sigman said, is that by the Spring of 2011, common assessment administration procedures and a common set of item release policies, a test security policy, and common definition of an English learner will be in place across all 26 PARCC states. Each consortium state would then be expected to adopt a common set of college- and career-ready standards by December 2011. Field-testing would occur in 2012 and 2013, with full PARCC assessments underway in 2014.

Sigman explained that California may have to consider how to assess additional standards, if other states have added to the common core standards — which will require additional legislation. In fact, she said, each of the PARCC actions, from through-course components to speaking and listening components, will require legislation.

In closing, she emphasized that these standards are summative assessments, not formative ones. This is not to say that tools such as computer-based assessments would not be able to give teachers more information in a more timely manner — but that the purpose remains summative, not formative.

In terms of California's readiness for computer-based testing (an element of PARCC), Sigman noted that California's assessment vendor had conducted both a survey and in-person site visits, concluding that computer-based testing would be feasible, but expensive — what Sigman described as "a heavy lift." An Education Week study similarly ranked California in the lowest quintile in overall technology leadership among states, with an average of 5 students per computer in California (compared to 3.8 nationwide).

> Outreach and education with the community on the purposes and uses of different assessments is the responsibility of PTAs, principals and teachers.

This remains a challenge as California moves to computer-based, adaptive assessments. "It's the right thing to do," Sigman said, "but it take funding and resources, making it a big challenge for us."

Resources on the Common Core California Standards (CCCS) and PARCC

CCCS: http://www.scoe.net/castandards/index.html

Common Core: http://www.corestandards.org/

PARCC Proposal: http://www.floe.org/parcc/

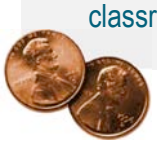
Oregon's Assessment of Knowledge and Skills (OAKS)

Tony Alpert, Oregon Department of Education

Tony Alpert, Director of Assessment for the Oregon Department of Education, joked that the acronym for Oregon's Assessment of Knowledge and Skills (OAKS) was appropriately pastoral and pleasant, just like Oregon itself.

Oregon's system is in fact far from the image of trees, especially as they relate to pencils and paper. It is computer-based, allowing for iterative assessments throughout the year. Students take tests online at three different points during the year, within an 8-month window from October to mid-May. The feedback is immediate — within 15 minutes — and is provided to teachers as well as aggregated for use by schools, districts, and the state. All of the 14,000 assessment items in the OAKS data bank were written by Oregon teachers; they also score the items.

> We need professional development on educational technology, not just having it in classrooms.



Alpert explained the rationale for computer-based testing. Although the initial investment in computers and test design was relatively high, the costs are reduced over the long term. In addition to providing instant feedback to both students and teachers, the system removes errors due to inter-class correlation by "testing the right student, with the right test, at the right time," Alpert said. Teachers also appreciate that testing can extend later in the year than pen-and-pencil tests, capturing more cumulative knowledge. And, "Oregon being Oregon," Alpert said, the carbon footprint implications of avoiding paper are another plus!

The computer-based testing allows the system to use adaptive testing, in which students receive tests questions differentiated according to the ability they demonstrate in answering previous questions. The technology also allows for constructed responses that offer much more information than single-choice items — and also offer opportunities for partial credit because students have shown more of their work.

Despite these many advantages, Alpert said, the system does introduce some new challenges related to access, fairness, bias, and interpretation. The tests also can be more time-consuming. Another issue is that as cognitive demand increases, so does the required language proficiency. A major challenge is how to increase cognitive demand for assessment purposes, without biasing English language learners and students with disabilities (e.g., for interactive items that may require fine motor skills).

Science Assessment in Minnesota

Dirk Mattson, Minnesota Department of Education

The Minnesota Comprehensive Assessments-Series II (MCA-II) were a response to testing requirements under No Child Left Behind; design work on the state's science tests began in 2004 as part of that response.

How do we get from where we are in California on assessments to the innovative assessments we've seen today?

The tests are based on scenarios. Students read a passage on a computer screen and then answer some questions about what they've just read. No paper is involved. Each scenario is designed to provide students with a realistic representation of classroom experiments and real-world phenomena, and to give students opportunities to consider science content at a higher cognitive level than would be possible with stand-alone items. Whenever possible, the scenarios take full advantage of the computer technology by inserting graphic animation, audio or video along with the written passages.

The scenarios are teacher-driven. Teachers did not want to rely exclusively on multiple choice tests, Mattson said, but instead wanted to develop something closer to the spirit and feeling of scientific inquiry. As a result, in addition to multiple choice items, the students' responses may include figural responses (in which students manipulate graphic elements or complete graphs and diagrams), short constructed responses, and extended ones.

In a demonstration of one of the scenarios, Mattson showed a video of a child riding a bike. Questions asked the student to identify elements on the screen — the bike to the park, the water bottle, an apple — in terms of energy changes (e.g., heat to mechanical, chemical to electrical, etc.). Next, as the cyclist continues on his way, the student is asked questions about his constant speed based on the trees he is passing; to chart data on how far he has traveled; and to explain what is happening to water molecules as they form droplets.

The conceptual design that began in 2004 led to field testing and a launch of the testing, with a new phase scheduled for 2011-2012. Along the way, the design and implementation teams encountered many challenges. These included crafting the RFPs in ways that would allow test publishers to recognize what they were after (which, after all, was new to everyone), determining a fair cost for these services (again, since there were no precedents), and funding the initial research and development efforts. A week before the testing window, Adobe released a new update of Flash, which was frustrating. Another discovery, Mattson said, was that "Kids click *everywhere*! As it turns out, there's no maximum, satisfactory level of quality control."

Alongside the challenges, Mattson reported, they were fortunate to have some key sources of support working in their favor. These included strong support from the Commissioner's office, enthusiasm and sustained commitment from Minnesota's science teacher community, and grant funding that allowed them to invest in the initial research and development.

Key lessons learned included convening a statewide Assessment Technology Work Group to define issues and pilot the software. This group became an important ally as the Department moved towards testing and deployment. Anecdotal reports of high levels of student engagement in the testing (even at the high school level) also were encouraging. Indeed, they experienced very few missing responses from these tests in the pilot — the fewest of any of the Department's tests.

Calibration of evaluators of "performance" assessment seems like a key element.

Mattson's advice to others considering similar scenario-based assessments included the following:

- Sin boldly. Sometimes, no one has learned the lesson before you.
- Give those designing content the message that they should shoot the moon, but "play where I can see you."
- Seek out "standardized flexibility," balancing the two.

In summary, scenario-based assessments can be both a blessing and a curse.

Reflections

David Pearson, Dean and Professor at the University of California, Berkeley, shared his reactions to the presentations on summative assessments and the models in other states.

First, he was struck by how the computer-based adaptive assessments not only offered greater efficiency, but also the promise of assessing performance closer to the student's real level of performance. The opportunities that technology offers for having students collect and analyze information — for complex cognitive tasks — is also important.

"I'm just thrilled to see this kind of work going on and really applaud our neighbors for this wonderful work," he said. California has been treated to a lot of great ideas for the future, he continued, and he hopes, given the political and budget environment, that the state will be able to act upon them.

Conclusion

Cal TAC's Chair and Vice-chair — Anne Marie Bergen and Barbara Shannon, respectively — thanked the speakers for their stimulating presentations and everyone for their participation.

Can we see more action ideas? Talk is great; action is needed.

As a group, Cal TAC members plan to reflect on the ideas generated during the symposium and draft a set of responsive next steps that can form the basis for the group's work over the next year, and beyond. To find out more and stay current with Cal TAC and its activities, please visit CCST's Web site: www.ccst.us.

Appendix A: Agenda

Working Toward an Assessment System with Value: Informing and Strengthening STEM Teaching and Learning

October 21, 2010
Sacramento Convention Center
1400 J Street, Room 307

8:30 A.M. **Continental Breakfast**

9:00 A.M. **Welcome, Introductions and Framing the Day** Anne Marie Bergen, Cal TAC
Karl Pister, CCST & CFTL

9:30 A.M. **Keynote Address**

Joan Herman, National Center for Research on Evaluation, Standards, and Student Testing

10:00 A.M. **A Focus on Innovative Formative Assessments** Lewis Chappelear, Cal TAC

Kathy DiRanna, WestEd's K-12 Alliance

David Foster, Silicon Valley Mathematics Initiative

Alastair Inman, Lexington Junior High School

Discussant: Sandy Dean, National Board Resource Center, Stanford University

12:00 P.M. **Lunch**

1:00 P.M. **A Focus on Innovative Summative Assessments** Diana Herrington, Cal TAC

Deb Sigman, California Department of Education

Tony Alpert, Oregon Department of Education

Dirk Mattson, Minnesota Department of Education

Discussant: P. David Pearson, University of California, Berkeley

2:30 P.M. **Synthesizing and Looking Ahead** Caleb Cheung, Cal TAC

3:00 P.M. **Adjournment** Barbara Shannon, Cal TAC

Appendix B: Participants

Presenters

Tony Alpert
Director of Assessment
Oregon Department of Education

Sandy Dean
Director, National Board Resource Center
Stanford University

Kathy DiRanna
Director
WestEd, K-12 Alliance

David Foster
Executive Director
Silicon Valley Mathematics Initiative

Joan Herman
Director, National Center for Research on Evaluation, Standards, and Student Testing

Alastair Inman
Science Teacher
Lexington Junior High School

Dirk Mattson
Director of Assessment and Testing
Minnesota Department of Education

David P. Pearson
Dean and Professor
University of California, Berkeley

Deb Sigman
Deputy Superintendent
Curriculum, Learning and Accountability Branch
California Department of Education

Attendees

Lynn Barakos
Professional Development Specialist
Lawrence Hall of Science

Kristin Nagy Catz
Assessment Specialist
Lawrence Hall of Science

Joan Easterday
Secretary
California Mathematics Project

Margaret Gaston
President
Center for the Future of Teaching and Learning (CFTL)

Lisa Giroux
Principal Consultant
Senate Office of Research
Beth Graybill
Principal Consultant
Senate Committee on Education

Susan Hackwood
Executive Director
California Council on Science and Technology (CCST)

Chris Hopper
Associate Dean of Professional Studies
Humboldt State University

Hui-Ju Huang
Associate Professor
Sacramento State University

Terry Janicki
Consultant
California Commission on Teacher Credentialing

Chelsea Kelley
Consultant
Assembly Education Committee

Attendees (cont.)

Julie Kidd
Program Officer
S.D Bechtel Jr. Foundation

Stacey Kyle
Research and Policy Associate
Center for the Future of Teaching and Learning

Phil La Fontaine
Director, Professional Development and Curriculum Support
California Department of Education

Nicole Lezin
Cole Communications

Brian Lim
Credentialing/ Assistant Professor
Sacramento State University

Maria Lopez-Freeman
Executive Director
California Science Project

Gretchen Miller
President
California Mathematics Council

Ann Miller
Director
California Science Project

George Miller
Senior Lecturer Emeritus, Chemistry
University of California, Irvine

Dale Oliver
Associate Dean of Natural Resources and Sciences
Humboldt State University

Rachel Perry
Director, Assessment, Accountability, and Awards Division
California Department of Education

Attendees (cont.)

Diana Rude
Consultant
California Council on Science & Technology (CCST)

Patty Scripter
Legislative Advocate
California State PTA

Suzan Solomon
Vice President for Education
California State PTA

Jane Stahler
Texas Instruments

Sheri Willebrand
President
California Mathematics Council

2010 California Teacher Advisory Council (Cal TAC) Members

Anne Marie Bergen, Cal TAC Chair
Teacher in Residence
College of Science and Math
Cal Poly, San Luis Obispo

Lewis Chappelear
Engineering Teacher
James Monroe High School

Caleb Cheung
Science Manager
Oakland Unified School District

Jeffrey Foote
6th Grade All Subjects Teacher
Kermit McKenzie Jr. High

Diana Herrington
Mathematics Teacher
Clovis High School

Stan Hitomi, Emeritus Member
Principal
Alamo Elementary School

Juliana Jones
Mathematics Teacher
Longfellow Middle School

Suzanne Nakashima
Elementary School Teacher
Lincrest Elementary School

Susan Pritchard
Science Teacher
Washington Middle School

Barbara Shannon, Cal TAC Vice Chair
Co-Director
Synergy Kinetic Academy

Brian Shay
Mathematics Teacher
Canyon Crest Academy

Katrina Williams
Elementary School Teacher
Harvest Elementary School