

October 7, 2005

MEMORANDUM

To: Participants of the 2005 MSP Evaluation Summit

From: Adding Value and Building Evaluation Capacity RETA Projects

Subject: Summary of the 2005 MSP Evaluation Summit

The MSP Evaluation Summit was held on September 15-17, 2005 at the Radisson University Hotel, Minneapolis, Minnesota. One hundred and twenty-seven participants included 40 MSP evaluators, 25 principal investigators and co-principal investigators, 5 NSF Project Officers, and 1 Department of Education representative. Advisory board members in attendance were James Altschuld, Audrey Champagne, Frank Davis, Arlen Gullickson, Frances Lawrenz, Donna Mertens, and Tom Romberg. In attendance from the Wisconsin Center for Education Research Adding Value RETA Project and Utah State's Building Evaluation Capacity RETA Project were Catherine Callow-Heusser, Heather Chapman, Scott Bates, Jim Dorward, Steve Lehman, Rob Meyer, Norman Webb, and Paula White. This memo summarizes progress made at the Summit.

Wednesday, September 15, 2005

Statistical Multilevel Models for Evaluating the Impact of Science, Engineering, and Mathematics Programs on Student Learning

The Evaluation Summit pre-session presented by Rob Meyer, University of Wisconsin-Madison, provided:

- An introduction to alternative statistical strategies for evaluating science, engineering, and mathematics programs and policies.
- Archetype evaluation models were presented to eliminate potential selection biases that may arise given different assignment/selection mechanisms such as unrestricted self-selection, matched-pair and random assignment.
- The evaluation models were presented within a unified framework that also included value-added models and models of student longitudinal data.

Evidence-Based Evaluation Findings Using Systems Change and Complexity Science Frameworks and Ways of Thinking

The Evaluation Summit pre-session presented by Michael Patton, Union Institute, focused on:

- Alternative evaluation designs and uses in support of more sophisticated matching of an evaluation to the systems nature of the project and situation that are the focus of the evaluation.

- Systems change efforts can benefit from systems modeling and systems change evaluation.
- Emergent situations can benefit from complexity understandings and rapid, continuous, real time monitoring approaches.
- Mixed methods approaches bring multiple perspectives to bear in delineating systems dynamics.
- Qualitative evaluation methods can be adapted to provide in-depth systems understanding and emergence monitoring applications.
- The workshop explored new directions in evaluation and their implications for evidence-based interpretations.

Thursday, September 16, 2005

Opening Plenary Session

Norman Webb, Principal Investigator, Wisconsin Center for Education Research and Cathy Callow-Heusser, Project Director, Building Evaluation Capacity in NSF's Mathematics and Science Partnerships, gave an introduction to conference participants by identifying the goals of the conference:

- Further community among MSP evaluators
- Address common issues
- Provide assistance in analyses
- Increase the knowledge of MSP evaluators about design, indicators, and conditions needed to successfully measure change in student learning over time
- Develop useful tools and designs for evaluators to attribute outcomes to MSP activities
- Apply techniques for analyzing the relationship between student achievement and MSP project activities to evaluate the success of MSP projects

Findings from the MSP Monitoring System

J. Frechtling (Westat), J. Hamos (NSF)

MSP knowledge Dissemination Project

D. Heck (Horizon Research), B. Miller (Education Development Center)

Breakout I

Strand 1: Student Learning and Participation

Beyond Achievement: Motivation-related Evidence from a Partnership Between a Targeted Project (TASEL-M) and a RETA (MSP-MSP)

A. Conley, S. Karabenick, J. Blazevski, J. Friedel, D. Pagni

The presenters reported on and explained:

- The development and initial administration of a scale to measure student motivation in mathematics courses
- The assessment tools, detail data collection efforts, and initial findings on student motivation
- The evolution of the feedback process, and how results from the student surveys are becoming part of TASEL-M's professional development

Assessing the Quality and Quantity of Student Discourse in Mathematics Classrooms

D. Weaver

Weaver presented on and summarized:

- The development of a student observation protocol to assess the quality of student's discourse regarding mathematics including efforts to increase inter-rater reliability, challenges to data collection, and findings from baseline data collection efforts
- The types of analyses that will be performed with the data and other research questions that can be addressed through these analyses
- Participants received a copy of the observation protocol used in the study

Using Data to "Make a Case" for Mathematics Reform Within a K-12 District

B. Sayler, J. Apaza, M. Austin

The presenters reported on and highlighted:

- The use of data to identify project needs and to inform stakeholders of progress.
- The most compelling findings thus far including a discussion of the venues and process used for sharing these data.

Special attention has been paid to gaps and patterns associated with Native American versus non-Native American students. Reduction of race-related disparities has represented a primary emphasis for the project. Preliminary findings have been generated through analysis of:

- Student performance on the state's multiple choice test
- Student performance on a free-response test developed by the Mathematics Assessment Resource Service
- Classroom observation ratings
- Student course-taking patterns
- Drop-out rates

Discussant: Norman Webb

Most significant ideas relevant to the MSP enterprise:

- Measuring motivation is an important outcome in addition to achievement

- Student discourse can be an important intermediary variable
- Stakeholders can benefit from having access to data. Need to unpackage the data to get at meaningful data

Methodologies used:

- Questionnaires on attitudes
- Observations of student discourse
- Multiple sources of data – observations, two assessments, professional development participation, descriptive data

Most important point(s) made during discussion:

- Student discourse and observation could be used as an intervention.

Issues raised that still need to be resolved:

- How to attribute student achievement growth to professional development?
- How many observations needed to obtain results? Reliability of observation data.

Strand 2: Teacher Change

PRISM Institute on the Teaching and Learning of Science and Mathematics
R. Henry

Henry discussed:

- The structure and content of statewide Institute meetings as well as regional institutes and disciplinary networks.

Comprehensive Evaluation of a Professional Development Program: Methods and Results in the PRMSP/AIACiMa Project
J. Arce, M. Bravo

The presenters focused on five dimensions of the evaluation of the PRMSP/AIACiMa professional development program:

- Participants' post activity reactions
- Participants' content matter and pedagogy learning
- Transfer of learning to the teachers' classroom
- School culture regarding support for this transfer
- Students' learning

Using Evidence from Teachers' Online Instructional Learning Logs in Evaluation and Decision-Making in an MSP

D. Heck, M. Wickwire, J. LaMaster

Three uses of evidence from the learning log data were examined in this presentation:

- The evaluation team uses evidence from the learning log data, in conjunction with other data collection from the evaluation, to support formative feedback to the partnership
- The Indiana University Mathematics Initiative partnership uses formative feedback and evidence from the logs to identify areas of common need and requests for support to refine its professional development offerings for teachers and to identify products that the partnership creates and disseminates
- The learning log data provide early evidence of impacts, and evidence for interpreting other data on impacts on teachers and their teaching

Discussant: Jim Altschuld

Strand 3: Institutes of Higher Education

PCMI and Districts Partner to Design Professional Development: Implementation Challenges and Evaluation Design

G. Burrill, N. Kher

The presenters discussed the challenges of measuring the efficacy of the PCMI professional development program:

- Given the small number of participants, traditional quantitative analysis doesn't produce robust statistical results
- To determine causality and directionality, longitudinal data are needed to assess change
- To understand the diffusion of new information in schools, important to learn who talks to whom about this information

Using Formative evidence and Formal Collaboration to Evaluation and Improve the Efficacy of an MSP/RETA Project

J. Labov, J. Garton, N. Shapiro, P. Maloney

The presenters discussed:

- The design of their formative assessments, how they were used, and data collected on ways in which participants have utilized the professional development workshops in the work of their own MSP project
- The planned collaboration between the National Academies and Change and Sustainability in Higher Education (CASHÉ) to undertake a more in-depth study

of the efficacy of this project with an emphasis on how these workshops might be influencing changes in higher education partners within the MSPs

Discussant: Judith Fonzi

Most significant ideas relevant to the MSP enterprise:

- Both papers remind us that we can pretty much count on the fact that our “perfect lesson plan” won’t be so perfect when the kids come in the door!
- In the case of PD3, teachers weren’t chomping at the bit to become lead teachers; classroom teachers weren’t signing up in droves to participate in the PD; and, the necessary curriculum data wasn’t as easy to get as they has assumed.
- In the Facilitating MSPs, project, workshop designs were not completely on the mark; and new, unanticipated workshops had to be added.

Methodologies used: The PD3 evaluation team is searching for some innovative tools for trying to capture the complexities of these MSP projects and they suggest Frank’s *Social Network Methodology*.

Most important points made during discussion: These papers can inform each other by probing what was not said. For example, probing Gail and Neelam about the role and work of the higher education faculty in their project could help inform Nancy and her team’s evaluation design.

Issues raised that still need to be resolved:

- How can the evaluation team be involved and informed at the level of detail AND in the timely manner that is often necessary to uncover the need for adjustments to the project and/or evaluation design?
- Who are you counting when you say, higher education faculty? How many, from what departments, what is their status (official and unofficial) in the university?
- What is the history regarding collaborations – within their respective departs, across university departments, within the community, with K-12 institutions or individuals?

Breakout II

Strand 1: Evidence-based Design

Coordinating Research and Evaluation Issues in an MSP Institute Partnership Project
S. Meyer, R. Heaton

The presenters discussed:

- The external evaluation conducted by RMC Research Corporation including a teacher survey that has been administered annually to project participants and a comparison group
- Project participants complete the Learning Mathematics for Teaching Survey of Content Knowledge for Teaching Middle School Mathematics
- The evaluation also includes interview with focus groups of teachers, school administrators, IHE faculty, and project staff; observation of project activities and classroom instruction, collection of student achievement data, and document analysis

Variation and Change in MSPnet's On-Line Community: Early Developments

J. Falk, B. Drayton, S. Lee, M. Ong

This presentation drew on:

- Quantitative and qualitative data to describe both emergent patterns, and their change and development over time.
- Quantitative data included web-usage data for MSPnet Hub and for project spaces. Qualitative analysis referred to types of contents and activities in project spaces. Analysis was presented on variation and use of the MSPnet from the point of view of the theory of communities of practiced.
- The presenters also explored evidence for effective practices for facilitation of discussion and information exchange within and between projects.
- Since the project is only 1.5 years old, the findings were provisional, and the presenters pointed to further research currently under way.

Building Upon Partner's Assets to Improve Mathematics Learning for Vermont Students

D. Harris, R. Quinn

This presenters focused on the themes of evidence-based design and changes in K-12 institutions to explore the ways in which the Vermont Mathematics Partnership accomplishes project goals while tailoring efforts to capitalize on partners' unique assets and to target specific needs:

- Teachers, and teachers in training, deeply understand mathematics and can translate their knowledge into high levels of student learning
- School support systems are rich with learning opportunities for students and teachers
- Partner schools and districts use valid and reliable ongoing assessments and feedback systems to continuously improve mathematics learning for all students
- Mathematicians and educators collaborate to develop high-quality professional development materials and protocols for teachers and teachers in training to build understanding of mathematics content, instructional practices, equity strategies and educational leadership

- Mathematicians and mathematics education faculty support collaborative research efforts among preK-12 educators, contributing to the state and national research base in the teaching and learning of mathematics

Discussant: Frances Lawrenz

Most significant ideas relevant to the MSP enterprise:

- The first paper was collaboration so that the research and evaluation components could work together and make both more efficient.
- The second was the MSPnet—who uses it and what they use. The big thing was that lots of public people use it.
- The third one was about designing unique interventions for each school—so the big deal would be that one size does not fit all.

Methodologies used:

- The first one just described what they were doing. Many different methodologies were used but results were not presented just how the research and evaluation worked together.
- The second one was counts of unique visitors and then of numbers of times for each visitor—so people would come and use it more often.

Most important points made during discussion:

- Using evidence or partnerships, partnerships within the project team
- School change and how needs assessment data can be used to have the program fit well

Issues raised that still need to be resolved: It was more descriptive and information oriented. The three presentations together emphasized communication, collaboration and use of information to design things.

Strand 2: Partnerships

Evaluating New versus Mature Partnerships: How Evaluation Questions May Change Based on Partnership Longevity

J. Frechtling, J. Winkler, J. Lara

The presentation illustrated:

- How their initial observations of the FOCUS partnership caused them to re-examine their definition of partnership in general.
- The tenure of the FOCUS partner collaboration appears to be a primary variable in how each element of a partnership definition is interpreted.

Organizational Mapping as a Tool for Understanding K-12 Partnerships to Improve Mathematics and Science Education

M. Clifford, S. Millar

The presentation addressed and discussed:

- How do leaders organize people and tools to accomplish partnership goals, as stated in the formal initiating documents?
- How does organizational mapping overcome faults of organizational charts made apparent in their work with SCALE, and strengths and weaknesses of the method.

Establishing Partnerships to Provide Evaluation Technical Assistance and Promote Evidence-Based Designs

H.J. Chapman, C.A. Callow-Heusser, J. Dorward

The Consortium for Building Evaluation Capacity (CBEC) is completing its third year as an NSF-funded Research, Evaluation, and Technical Assistance (RETA) project. Through use of the Network for Evaluation Technical Assistance (NETA) and a group of nationally recognized experts in the field of evaluation, the CBEC has offered:

- Evaluation technical assistance to a number of MSPs with the overall goals of improving evaluation quality and increasing evaluation capacity for the MSPs.
- The challenges the CBEC has faced in offering and providing technical assistance to projects have led to a number of suggestions for improvement to the CBEC's original model of technical assistance.

Discussant: Frank Davis

Most significant ideas relevant to the MSP enterprise:

- The Frechtling, Winkler, and Lara paper advanced several ideas about indicators measuring activities that characterize partnerships, including the existence or understanding of shared goals, intellectual engagement and understood obligations to work toward shared visions.
- The Clifford and Millar paper offered some new ideas on how to look at participation in partnerships. Mapping the working relationships of various individuals, in terms of previous institutional relationships, as well as program goals, offers the possibility of evaluating the dynamic and fluid work of partnerships.
- The Chapman, Callow-Heusser, and Dorward paper offered a look at the problems of providing evaluation technical assistance – both from the point of view of clients and evaluators. The paper noted the problems of shared understanding of goals of MSP work, both from the perspective of clients and evaluators.

Methodologies used: All of these papers should be characterized as work in developing grounded theory. They all explored issues of defining what partnership may mean through observations/explorations of on-going partnership experiences. They all posed questions/hypotheses about the nature of partnerships that need further exploration.

Most important points made during discussion:

- All the points made illustrated the need for much more work in understanding partnerships, both conceptually and empirically.

Strand 3: Teacher Change

The Effects of Professional Development on Improving Mathematics and Science Instruction (MSP PSD Study)

R. Blank

Blank summarized progress after two years of a three-year study under the MSP RETA program:

- The study is being conducted by a research team comprised of staff from the Council of Chief State School Officers, American Institutes for Research, and the Wisconsin Center for Education Research.
- One purpose of the MSP PD study is to test the use of teacher self-report survey and web-based tools for collecting, analyzing and reporting data on the quality of professional development, and the usefulness of these data tools for evaluating effects of professional development.
- The work is being conducted in four MSP sites from the first cohort of MSP grant sites (starting Fall 2002): two comprehensive projects and two targeted projects.

Evaluating Professional Learning Communities Using Mixed, Quantitative and Qualitative Methodology

J. Monsaas, M.J. McGee-Brown

This presentation reported on:

- The database set up to track the participation of the Learning Communities (LCs) in the PRISM project.
- Quantitative data have been gathered on teachers' use of inquiry-based teaching and learning Strategies (ITAL) in both PRISM and non-PRISM schools.
- The comparative results were provided on teachers' use of inquiry based teaching and learning practices as well as student improvement in science and mathematics achievement.
- Extensive qualitative data have also been gathered and preliminary findings were presented on the effectiveness of the LCs and teachers' perceptions of how they are influencing teaching and learning.

Discussant: Donna Mertens

Most significant ideas relevant to the MSP enterprise:

- Blank: Professional development – reflection of research-based definitions of quality.
- Monsaas and Brown: Evaluation of learning communities – learner-centered.

Methodologies used: High correlation; cross-validation.

Most important points made during discussion:

- Is there a representation of underrepresented groups in the research-based evidence used as criteria for effective practice?
- Are invitations issued to members of under-represented groups to participate in decisions about the evaluation and intervention?

Issues raised that still need to be resolved:

- Test the assumption: if you set challenging content standards for all students, will that improve student performances? Will presentation of the same content to the majority and minority students close the achievement gap?
- Effect of attrition on outcomes?
- We results regarding quality of professional development shared with all four projects?
- Criteria did not include reference to adaptations based on cultural diversity.
- How important are the criteria for selecting experiment versus control groups other than participation versus non-participation.

Breakout III

Strand 1: Student Learning and Participation

Measuring the Effect of the Milwaukee Mathematics Partnership on Student Achievement

C.M. Walker, J. Gosz, D. Huinker

The presenters addressed the following questions by using two years of achievement data and linking it to variables measured by a self-report instrument:

- Is there a greater increase in student achievement in mathematics for schools with collaborative and supportive Learning Teams
- Are mathematics teachers in schools with collaborative and supportive Learning Teams more likely to report teaching mathematics in a manner aligned with the NCTM Standards and does this result in a greater increase in student performance?

Using Participation Maps in the Evaluation of Participation

J. Watson

To address the challenges of large interventions projects, a system was developed for mapping the number of SCALE interventions across each district, grade and content area:

- The system utilized a combination of database and spreadsheet manipulations to create quantitative and graphical representations of levels of SCALE activities.
- The system provides SCALE with a platform for developing a systematic approach for determining areas of high impact.
- The system also provides a decision support to SCALE leadership and management and informs the evaluation work.

Integrating Curriculum Guides, Quarterly Benchmark Assessments, and Professional Development to Improve Student Learning in Mathematics

K. Hyde, V. Mann, C. Manrique, T. Shanahan

Teachers in a large urban school district in southern California used standards-based curriculum guides, quarterly benchmark assessments, and focused professional development to improve the achievement of their students in math content as measured by the California State Test in mathematics and reported as part of each school's Annual yearly progress.

- The findings indicated that the average growth of the students in mathematics far out-paced their growth in English Language Arts as well as the average growth for students in the county and the state in mathematics.
- The findings also supported a three—pronged approach consisting of curriculum guides that organize and pace the content standards and include model tasks for student outcomes supported by standards-based quarterly benchmark assessments that give the teachers timely feedback about student conceptual understanding of the standards.

Discussant: Jim Dorward

Most significant ideas relevant to the MSP enterprise:

- The Milwaukee project examined a model of creating school leaders as an impetus for systematic change
- Watson's SCALE study developed a system for mapping SCALE interventions across each grade and district
- The Irvine Math Project measured the impact of focused professional development on improving student performance.

Methodologies used:

- The Milwaukee project created a teacher self-report survey instrument to measure professional development based on the goals and objectives of the partnership and tried to capture school level variability.
- The SCALE mapping project used a combination of database and spreadsheet manipulations to represent levels of SCALE influence.
- The Irvine Project used a three-pronged math intervention combining implementation of curriculum guides, quarterly assessments, and professional development.

Most important points made during discussion:

- Milwaukee project linked two years of student achievement data to variables measured by teacher self-report.
- The Irvine Project found that combined implementation of curriculum guides, quarterly assessments, and professional development led to a twelve percent gain in student achievement.

Issues raised that still need to be resolved:

- Milwaukee study had to examine whether the information the teaching specialists reported was actually representative of what occurred in the schools, examined the effect of missing data, no way to link teacher responses to their classrooms.
- Irvine Study had high teacher turnover presenting problems in successful implementation and evaluation, plus teachers self-selected to participate.
- SCALE mapping project faced challenges of weighting activities and tracking data sources.

Strand 2: Teacher Change

Teacher Change in High School Science: Findings from the First Three Years of the Vertically Integrated Partnership (VIP) K-16

K. Raue, J. Frechtling, X. Zhang, B. Hedges

Findings from the first three years of implementation of the Vertically Integrated Partnership (VIP) K-16 were discussed. Data were drawn from two sources:

- Pre- and post-program surveys of VIP's Biology Cohort
- Classroom observations of a sample of VIP's Matter and Energy/Earth Space Systems Cohort.

Changes in teachers' instructional practices were examined, including their use of:

- Inquiry activities
- Assessments

- Strategies to promote reading and writing in science
- Teachers' beliefs about teaching

Comparison of Inservice Elementary and Middle School Teachers' Understanding of Selected Light Concepts

R. Atwood, J. Christopher, R. McNall

Three institutes serving 72 elementary teachers and 51 middle school teachers were offered at different sites across the AMSP target region. This presentation addressed the research question:

- How does the performance of inservice elementary and middle school teachers compare on selected standards-based light concepts before and after instruction designed to address the concepts?

Case Studies to Evaluate Teachers' Transfer to Classrooms of Learning Derived from a Professional Development Program: Methods and Results in the PRMSP/AIACiMa Project

To study the transfer to classrooms of learning derived from the Professional Development Program, five case studies were carried out in the evaluation of the PRMSP/AIACiMa Project. The following questions were addressed:

- How have the project's participants transferred what they have learned in the AIACiMa Professional Development to their classrooms?
- How are they integrating assessment to the teaching and learning processes?
- What learning with understanding has been evoked in the students by the pedagogical practices employed?

Discussant: Rosalie Torres

Most significant ideas relevant to the MSP enterprise:

- In-depth case studies with school as unit of analysis resulted in rich description used to help revise staff development.
- Pretest of teacher knowledge on selected concepts showed little difference between elementary and middle school.

Methodologies used:

- Case studies – in-depth
- Classroom observation
- Survey of teachers regarding classroom practices
- Teacher content knowledge assessment

Most important points made during discussion:

- Teacher change is influenced by teachers' prior knowledge and attitudes as well as testing mandates in schools and pressure to cover prescribed curriculum.
- Teachers are making some progress in using new classroom practices, but not as much as hope for or expected.
- Case study approach and reporting of qualitative findings provide convincing evidence of "what is happening in classrooms."
- Multiple sources of data (observation, teacher self-report, student interviews, or survey about classroom practices) tend to converse/corroborate each other.

Issues raised that still need to be resolved:

- How to motivate teachers to participate in data collection activities? Burden is significant with all evaluation activities for MSPs.
- How to accurately assess teachers' conceptual knowledge in math and science.

Strand 3: Evidence-based Design

Exploring the Relationship Between the Use of Standards-Based Instructional Materials and Student Achievement in Mathematics: A Pilot Study

J. Apaza, B. Sayler, M. Austin

As part of this project, a pilot study is underway to explore the relationship between the use of standards-based instructional materials and student achievement.

- Two years into the pilot, twenty-three fourth grade teachers from Rapid City have been involved, and each was surveyed about the degree of implementation of the district's newly adopted, standards-based instructional materials, Investigations in Number in Data, Number and Space.
- Approximately 500 students in the classrooms of the selected teachers were given both the Dakota Student Test of Educational Programs (DSTEP) and the Balanced Assessment in Mathematics.
- To examine the impact of instructional materials on student achievement, results from the two assessments were compared with self-report data from teachers about the degree of implementation of Investigations.
- Scores on two assessments were examined in conjunction with teachers' self-reports of the percentage of time they spent using the district's recently adopted, standards-based instructional materials for mathematics.

Triangulating on Curriculum to Motivate and Inform Systemic Curricular Reform
R.T. Housang, L.S. Cogan, J. Ferrini-Mundy, W. H. Schmidt

This presentation examined the efforts of one MSP to stimulate systemic curriculum reform through evidence-based activities that provide a comprehensive portrait of a school district's mathematics and science curriculum.

- To measure curriculum at the district (Intended) level, curriculum experts completed a district 'Road Map.'
- To complete this instrument, experts indicated at which grade(s) a topic was intended to be taught according to the district's mathematics and science curriculum standards. Results from the student assessments (Attained) were reported according to TIMSS Curriculum Frameworks categories.
- This presentation included examples for the mathematics reports generated for districts and examples of how this type of multi-faceted curriculum portrait has been used to focus and refine districts' curricular expectations and practices.

Discussant: Tom Romberg/Catherine Callow-Heusser

Most significant ideas relevant to the MSP enterprise:

- The U-shaped scatter plot relation (use of standards-based materials to gain scores)
- The use of TIMSS methodology to gather data in 62 school districts

Methodologies used:

- Correlations between various measures
- Matrix sampling, adapted surveys from TIMSS

Most important points made during discussion:

- Two types of teachers
- Exploratory evidence
- Building a database

Issues raised that still need to be resolved:

- Need for data on quality of instruction
- Usefulness of the data summaries

Friday, September 16, 2005

Opening Plenary Session

Finding Value and Meaning in the Concept of Partnership

G. Kingsley, M. Waschak

Kingsley and Waschak explored the following partnership issues:

- The relationship of three distinctive conceptualizations of partnership with the factors that respondents identified as critical to the operations of partnerships and the likelihood of achieving outcomes and impacts.
- How the variance in the conceptualizations have challenged and changed the way in which their RETA project is approaching the task of modeling partnership behavior that they will use in their case study.

Breakout IV

Strand 1: Evidence-based Design

Evidence-based Design from the Mathematical ACTS MSP project at the University of California-Riverside

K. Bocian, R. Torres

In this presentation, the authors described the following evidence-based design issues related to the project evaluation:

- The project's decision to create an observation instrument rather than use one of several existing instruments
- The collaborative development process involving core team members
- Training of observers
- Procedures for establishing inter-rater reliability
- The use of the instruments
- Analyses conducted
- Lessons learned during the pilot year

Aspects of Minority Student Retention in STEM Disciplines and Evaluation of a Major Retention Program

J. Altschuld, Y. Lee, J. White

The context of the evaluation was the Ohio Science and Engineering Alliance (OSEA), a consortium of fifteen universities in Ohio funded via the Louis Stokes Alliances for Minority Participation (LSAMP) conducted under the aegis of NSF.

- The evaluators reviewed the literature for key variables and evaluation and research studies related to the persistence of underrepresented minorities in STEM disciplines.
- Highlights of the literature and their influence on the evaluation of OSEA was described, followed by the design of the evaluation strategy and its instrumentation. The rest of the session focused on a summary of evaluation results and possible implications.

Using Evaluation as a Bridge in Partnership Development

C. Tananis, J. Pane, N. Bunt, S. Olmstead, V. Williams

In this presentation, the authors described the structure and function of the integrated approach of the evaluation of the Southwest Pennsylvania MSP.

- Examples were provided of how evidence has been used to drive decision-making and strengthen partnerships across stakeholder groups, and the presenters discussed dilemmas evaluators face in conducting a more participative form of evaluation within a complex partnership such as the MSP.

Discussant: Arlen Gullickson

Strand 2: Changes in K-12 Institutions

Leading and Planning a Professional Development Program

C. Copolo

Copolo examined the effects of the TASC professional development plans on teacher and student growth in four participating school systems. The TASC program leases NSF-endorsed curriculum kits and trains teachers using these kits with partner school systems. These systems developed various plans for training their teachers and for leasing these kits from TASC. Evaluation of the impact of the TASC program focused on:

- Participants' comfort with teaching science
- Changes in teacher attitude toward science investigation using the Test of Science Related Attitudes (TOSRA)
- Perceived changes in classroom practices
- Observed use of kits as well as documented use of journals with kit activities

Student impact measurements included changes in student attitude toward science investigation using the same TOSRA instrument.

Thinking About Mathematics Instruction: A Preliminary Investigation of Mathematics Leadership content Knowledge of Principals in Ten MSP Sites

B. Nelson, G. Jonson, K. Reed

The presenters described baseline demographic characteristics of 400 elementary and middle school administrators who participated in the first phase of a five year research study, Thinking About Mathematics Instruction (TMI). These administrators were affiliated with ten Mathematics and Science Partnerships. In addition, two dimensions of the administrators' mathematics leadership content knowledge (LCK) were discussed, including:

- Mathematics content knowledge
- Beliefs about the nature of teaching and learning

Assessment of Student Learning With Understanding: Evaluation and Professional Development Activities in the PRMSP/AIACiMa Project

M. Aguierre, M. Bravo

The promotion of students' learning with understanding of mathematics and science content is a main aim of the PRMSP/AIACiMa Project.

- Its assessment is a major endeavor of the Project's assessment and evaluation teams.
- Instruments, procedures, and results from these assessment processes were presented and their implications for both evaluation and professional development were discussed.

Discussant: Jay Labov

Most significant ideas relevant to the MSP enterprise:

- Increased use of science kits, confidence in use (inquiry-based), more writing, less use of textbooks
- What is leadership content knowledge in math? Does it change with levels of learning and training?
- Evaluation/scoring of assessments has become a direct intervention in professional development

Methodologies used:

- Pre/post attitudes
- Pre/post random assignment, survey, and knowledge assessment
- Assess lesson plans, student assessments, and surveys

Most important points made during discussion:

- Teachers become mechanical in use of kits
- Content/conceptual knowledge is low, trying to change to reform-oriented thinking
- Professional development now includes evaluation/scoring of assessments because of how much they learned from it

Closing Plenary Synthesis and Discussion

Through open discussion during the closing plenary, participants identified fifteen unique ideas raised that were raised over the course of the Evaluation Summit:

1. Partnerships are complex beasts with multiple dimensions and not one concise definition and evaluating has own level of complexity, but need to look at because not just about program and outcomes but impact of partnerships.
2. Evaluation portfolio needs to be diversified, non-linear, and developmental. Go home and incorporate some of what we've learned here into our design.
3. We have to keep our goal very clear but have to keep adapting constantly from data we get from evaluation. By making changes, we are being consistent with what we want to do—keep make adjustments to reach goals, keep eye on the prize but have to make adjustments because evaluation informs.
4. Evidence comes in both empirical and wisdom of practice forms, value both, use in our projects to build plausible arguments for what we do. Use two kinds of evidence to build arguments.
5. Heard a lot about different instruments and emphasis should be on richness instruments, broadness, think about as a tool set, instrumentation is a form of knowledge building.
6. Lessons learned, became very clear in this meeting that instead of resisting when lessons learned are not what we want, we really have to look at what the evaluation data is telling us. Understand what data are staying.
7. Challenges of random assignment, got ideas about delaying treatment, telling both treatment and control precisely what evaluators will be looking for (instrumentation as intervention). Make sure everyone has same base.
8. Felt that it will be difficult to attribute change to MSP activities—a lot of presentations, seems will be a challenge we all will have. Attribution of change to MSP activities complex and difficult, but people are trying to focus on and measure this. This program has more emphasis on measuring change.

9. Comprehensive MSPs, brought cadre of members from MSP—they were able to take data from investigation and use in a timely way so people could use it and use effectively.
10. Talked a lot about complexity of partnerships, but have many attributes. Talked about difficulty of randomized design. Are there particular attributes of partnerships that lend themselves to better data collection and analysis and design? Efficiency.
11. NSF MSP model has different lenses, projects, RETAs, MIS, knowledge management. All working together and all part of larger piece and how do they operate in partnership.
12. Findings, when first cohort met for first time, only allowed to use two overheads but all about design. Systemic got to end what are common findings? Wish we had presented findings first, design later. Should think about those two transparencies.
13. One of the practical aspects of this whole thing is defining the interface for exchange of information. It's not as intuitive as it might sound, establishing interface for exchange of information, continued examination of interface.
14. Surprised by number of external evaluators who participated with project staff to do presentation. Are we building capacity of MSPs? Would have expected more project people to be doing presentations. Development of capacity to do evaluation?
15. A lot of tools being developed, models, etc., want to comment that the evaluation work of MSPs is a tremendous source of knowledge generation. Coming through creative evaluation work that is underway.

The following ten evaluation issues and ideas were raised by discussants:

1. Amount of data being collected. Efficiency—can't use it all. Are we collecting the exact data, the data that is most important, is it essential? Relates to design.
2. Utilization of data. In some cases, data very useful. Giving insights that have never had before, productive, raising questions. In other cases, so much data that how will it be used? Who will use it, report it?
3. Aggregation of data across site, across sites (meta-analysis). Each of partnerships so unique, configurations, instruments, few effect sizes reported. If trying to look at and aggregate data might be difficult.
4. What is good enough? All struggling with this. Attribution question, what is good enough and what do you compare it with? Difficult issue.
5. Instrumentation—valuable. How this can feed into implementation. Use instrumentation, vary process to get to goal. How do you take instrumentation, data, feed into implementation? Do evaluators hold information that would be useful to project?

6. Formative evaluation, how the role of formative evaluation plays.
7. Did not see as much as we thought we might see of drilling down. Defined, controlled studies rather than doing whole ball of wax. This notion of time, timeliness, sometimes easier to come together and do small study on a timely basis. Drill down, get info so can actually see what is happening.
8. Ways of categorizing projects into two or three groups. Focus on professional development. Commonality. Simplicity. Three or four different kinds of evaluation models—summative, formative, developmental, input, output.
9. Participant burden. Smaller districts where continually ask people for information, may be a problem.
10. We're very encouraged. Phenomenal, amount of effort, energy. Amount of work, data, findings. When do you expect to see results? Horizontal horizon. When do you expect to see results of intervention. Two years minimum, three or four years. Some who had findings only had a little bit—but maybe after two years, it's pretty good.