MSP Evaluation Summit Participant Evaluation Summary
September 15-17, 2005

On September 15-17, 2005 at the Radisson University Hotel in Minneapolis, Minnesota, 127 participants met for the MSP Evaluation Summit sponsored by the Adding Value to the Mathematics and Science Partnership Evaluations and the Building Evaluation Capacity Research, Evaluation, and Technical Assistance (RETA) projects, funded by the National Science Foundation. Participants included MSP evaluators, principal investigators, and project staff. Advisory board members in attendance were James Altschuld, Frank Davis, Audrey Champagne, 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, Steve Lehman, Scott Bates, Jim Dorward, Rob Meyer, Norman Webb, and Paula White.

The Evaluation Summit covered important aspects of evaluation with a focus on the following five themes: student learning and participation, teacher change, institutes of higher education, evidence-based design, and partnerships. The Summit began with two concurrent pre-sessions on September 15 including one on statistical models and HLM for the sciences, engineering, and mathematics by Rob Meyer, an economist and statistician from the Wisconsin Center for Education Research and a member of the Adding Value RETA project. The other pre-session was on using evidence to effect systems change in complex organizations and was facilitated by Michael Quinn Patton, a leading evaluation authority and author of several books on evaluation design, from the Union Institute.

The two-day Summit on September 16 and 17 was organized into three plenary sessions and eleven paper-reporting sessions related to the research or evaluation of the MSPs. The paper sessions were organized around five major themes: student learning and participation, teacher change, institutes of higher education, evidence-based design, and partnerships. The specifics of these topics were elicited from the MSP researchers and evaluators. Each session included discussants who were responsible for reacting to and critiquing the papers. The conference ended with a closing plenary synthesis and discussion summary on specific evaluation challenges and identifying procedures for continued work. The Evaluation Summit summary is available on the Adding Value Web site under Conferences/Meeting 6/Meeting Minutes: http://www.addingvalue.org/Conferences/Meeting%206/Evaluation_Summit_Summary_Sept_15-17_20051.pdf The discussion and networking opportunities aided the Summit in achieving its overall goal of clarifying evaluation needs through networking and intellectually rich conversations among MSP evaluators and RETA representatives.

Summary of Responses

What follows is a summary of the evaluations completed by the Evaluation Summit participants. The response rate was 40 percent, with fifty-one of the one hundred and

twenty-seven MSP, RETA, and Advisory Board participants completing evaluations (not all participants responded to all questions). Nearly all participants' comments are provided below, except in cases where comments were repetitive or illegible. Participants were asked to respond to each of the following aspects of the conference:

The pre-conference sessions
The Evaluation Summit sessions that you valued the most
The Evaluation Summit sessions that you found the least beneficial
How closely the Evaluation Summit sessions matched expectations
Large group sharing of projects
Break-out sessions
Other opportunities for networking
The Evaluation Summit overall

Pre-Session Response Summary

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

Participants were asked to rate each of these items on a scale from 1 to 5, with 1 signifying the lowest rating, "None" or "Not at all" and 5 signifying the highest rating, "a great deal."

How much did you gain from the Evaluation Summit pre-session?

The pre-session by Rob Meyer had a total of 17 participants and a 82 percent response rate to the participant evaluation. On a scale of one to five, the average rating for gain from this pre-session was 3.3 (some-what), with 14 percent of respondents giving a rating of 2, 21 percent of respondents giving a rating of 3, 36 percent giving a rating of 4, and 36 percent giving a rating of 5.

How closely did the pre-session match what you expected?

In terms of matching expectations, the average rating for this pre-session was 4.8 (nearly a great deal), with 29 percent of respondents giving a rating of 3, 36 percent giving a rating of 4, and 36 percent giving a rating of 5.

How useful did you find the pre-session overall?

In terms of how useful participants found this pre-session overall, the average rating for this pre-session was 4.0 (a lot), with 14 percent giving a rating of 2, 29 percent of respondents giving a rating of 3, 21 percent giving a rating of 4, and 43 percent giving a rating of 5.

Participants provided the following comments on the pre-session by Rob Meyer:

Got lots of good ideas that I want to try out.

Excellent presentation and content.

Clear and engaging workshop.

Thanks. My background on HLM is low. I learned lots, game me issues we need to consider for our MSP.

Nice introduction to value-added models and consideration of selectivity issues. Would like to have seen more detailed examples of how to get up and run analysis models.

A bit more emphasis on actual data analysis would be helpful. Some review of output was included, but resultant SAS code used to generate this output would have been very helpful for someone trying to learn HLM as a new method of analysis. Instructor was very competent and highly energetic! There was a lot of attention given to the what and the why and the results of an analysis, but more "how to" would have been really beneficial to newcomers to value-added models. What was covered was very good – my comments relate to additional content.

Evidence-Based Evaluation Findings Using Systems Change and Complexity Science Frameworks and Ways of Thinking, **Michael Patton**

How much did you gain from the Evaluation Summit pre-session?

The pre-session by Michael Patton had a total of 36 participants and a 78 percent response rate to the participant evaluation. On a scale of one to five, the average rating for gain from this pre-session was 4.2 (a lot), with 7 percent of respondents giving a rating of 2, 11 percent of respondents giving a rating of 3, 32 percent giving a rating of 4, and 54 percent giving a rating of 5.

How closely did the pre-session match what you expected?

In terms of matching expectations, the average rating for this pre-session was 4.0 9 (a lot), with 29 percent of respondents giving a rating of 3, 38 percent giving a rating of 4, and 33 percent giving a rating of 5. Four respondents indicated that they had no prior expectations.

How useful did you find the pre-session overall?

In terms of how useful participants found this pre-session overall, the average rating for this pre-session was 4.2 (a lot), with 14 percent giving a rating of 3, 32 percent giving a rating of 4, and 54 percent giving a rating of 5.

Participants provided the following comments on the pre-session by Michael Patton:

Although I didn't gain much "knowledge", the experience of participant questions, and sharing questions about how a MSP might use this approach, was informative. The strategies presented in this workshop are critical to effective evaluation of NSF projects, and other educational projects. It would be quite useful to have Patton give a more focused presentation to NSF staff so that they would understand and expect this component in overall NSF evaluation plans.

Transformative – inspired. Thank you! Having "permission" to engage in linear/systems and emersion evaluation all at the same time is freeing.

This workshop provided me with a thought-provoking framework to examine my MSP. Patton, like his topic, was dynamic, providing insightful examples to questions that emerged.

Thank you for inviting Dr. Patton who's a legend in qualitative research.

Patton is a dynamic speaker and bases his research on his actual current experiences which is useful and refreshing.

Thought-provoking. Highly informative. Intellectually engaging – for the whole day! Evaluation has to be flexible, not rigid!

As a novice, this is quite useful. It would helpful to give the theoretical framework and how to implement and have a few less (or shorter) "stories"/cases. For some, it seemed the point was to talk about his accomplishments (e.g., 5 programs funded, one not continued (his) but his was the only one that turned out to be truly important. Remember the audience.

Helped to define and elaborate on multiple evaluation designs to meet different programmatic outcomes.

Very timely for me despite the countervailing demands from NSF. A challenge to meet the opportunities inherent in his workshops but the benefits could be huge.

Excellent workshop. Very relevant to some of the "real world' issues of MSP evaluation.

This workshop helped to give language to describe much of what I have found perplexing when trying to use a logic model to describe our program. The articulation of non-linear models opens new ways to look at our project and use evaluation products.

This helped me greatly in thinking about how our large and complex MSP should be evaluated. It will be a systems approach.

This session was an intellectual pleasure. I appreciate Michaels' critical viewpoints and humor. He doesn't shy away from the complexity and messiness of the work – creating a stimulating environment. The day flew by and was packed with much that I will revisit and mull over and hopefully incorporate into our work.

Patton's extensive experience and examples were wonderful. He kept me interested all day! However, I doubt I'll be able to remember much – there was so little opportunity to reflect on how these ideas affect me and my project – a structured break to talk or write might have helped me along the way – I thought there was going to be an "applications" piece this afternoon. I would have appreciated that.

It was fabulous. I learned a lot. I only wish there was time to do the group work – more applications of systems to linear evaluations on our own work.

The workshop was excellent (significant content, well presented, engaged audience) until around 2:30pm. It seemed at that point, to move to individual questions that led to the presentation taking from his rich experience, but less connected. I was also disappointed that network theory wasn't developed.

I would have liked more specific examples of how to design and carry out system and emergent evaluation.

Too much conference. Active learning is needed to construct knowledge.

A few too many stores; too little time trying to apply concepts to participants' own projects. Quite thought-provoking and a very useful perspective (with wealth of experience) nonetheless.

Michael did a good job of presenting the conceptual framework of complexity science and systems change. Would've loved more time to do some work on our own MSP in small groups.

Which Evaluation Summit sessions did you value the most? Explain why.

Respondents identified a variety of sessions that they valued the most. Of the 51 evaluation responses, 14 participants identified the partnership session by Gordon Kingsley as the session that they valued the most, several respondents identified the opening plenary sessions on both the first and second days, several identified the pre-sessions by Meyer and Patton, the session on the Puerto Rico MSP, the session on assessment of student learning, evaluation of student learning and teacher change, evidence-based design, as well as the closing session.

The following general comments were made regarding the Evaluation Summit sessions that participants found the most valuable:

The opportunities to hear about specific evaluation instruments and results and cases in which this data is directly used by districts and partners.

All the sessions I went to. They were all very enriching, sharing knowledge from different sources.

The topics were relevant to all MSP structures.

All sessions contributed to my thinking. Discussants were provocative and participants' questions were important to both presenters and the audience. The structure and design worked well to establish the environment for serious exchanges.

The sessions that discussed methodology more than results were most useful to me.

I came away with ideas I can use in our own MSP Evaluation.

All of the various methodological discussions were very helpful for looking back at our own work and our methods.

On the pre-session by Michael Patton, participants commented:

I learned about the importance of looking at the whole picture and the need to take it into consideration to get results.

The pre-session by Patton was important for validating the emergent design. I wish his talk was available on tape for the rest of our team to see!

Patton's pre-session provided legitimization for honoring complexity/chaos in the evaluation process as well as the MSP project itself.

Patton's pre-session was engaging and scholarly, it provided a very helpful framework for looking at evaluation.

Patton's systems approach was particularly valuable. It was both recognition of reality and working outside of the box. Also, of all the sessions, it best allowed and encouraged participant involvement.

On the pre-session by Rob Meyer, a participant commented:

I developed ideas that I can apply to the quantitative part of our MSP evaluation; I got specific techniques.

On the session on partnerships, participants provided the following comments:

The partnership discussions opened my thinking to new possibilities.

The Friday morning session on the nature of partnerships got to the essence of what the MSPs need to think about: what they are and what they can and should be.

The partnership workshop provided a framework for linear and emergent designs.

Gordon Kingsley's presentation on partnerships was an excellent big picture view that provided not one way to look at the issue, but a set of ways that are quite useful.
Finding value and meaning in the concept of partnerships – interesting to identify our form of partnership and note that coordination and collaboration are different.

Kingsley's session on partnerships showed how fuzzy the notion of partnerships is and how inadequate a logic model is in doing more than getting started.

Several participants indicated that preparing papers and giving their presentations served as a useful tool in their own MSP work. Participants commented:

Our presentations – we got good feedback that was helpful.

Our own presentation was a great "process" for the team. I agree this should be an on-going "summit."

Networking was outstanding and so was input for our own project generated through presentations by our own team and others. Gained a lot in preparing for the Summit as well.

On the session on the Puerto Rico MSP, participants commented:

They presented interesting data that are relevant to the project I evaluate.

The work from Puerto Rico because they brought in the cultural context and modifications to research-based evidence to be responsive to under-represented groups.

On the summary plenary session, participants commented:

The wrap-up. The key ideas from the Summit were identified.

The final plenary helped consolidate experiences and gains.

The last plenary session was excellent. It's the best closing I've ever seen across all conferences in the last 14 years (science and education disciplines). Excellent enactment of practical wisdom, modeling constructivism and best evaluation and classroom practices and it was FUN! What learning is all about!

Which conference sessions did you find least beneficial? Explain why.

As far as identifying sessions that participants found the least beneficial, some participants indicated that they found all sessions to be beneficial. For example, participants commented:

None, the variety and quality was very stimulating!

All were helpful on some level.

All were useful to me.

Close to half of the participants provided an example of a session or an aspect of the conference sessions that they found least beneficial. Below are some of the comments:

Those that simply reported findings.

The break-out sessions that only explained the project design. The designs are mostly the same.

The sessions where people tried to explain their whole project, rather than trying to just highlight "one slice." Explaining the whole project resulted in rushed generalities that weren't specific enough to be useful.

The partnership discussion, but I am not involved in that aspect with my RETA.

Those on global evaluation of the MSP program as a whole. This doesn't mean I don't think such evaluation is important, it reflects that this is less relevant to the goals of our own project in particular.

The break-out sessions varied greatly in quality. The length of time didn't seem long enough to powerfully use the discussant role.

Several of the presentations in the break-out sessions were not focused or well constructed for the time available in the conference, which made some of these sessions less beneficial than they might have been.

Disappointed by the MIS results. Not very interesting results. Made it that much more frustrating given the Herculean efforts that went into entering the data last fall and winter.

Participants were asked to rate each of the following items on a scale from 1 to 5, with 1 signifying the lowest rating, "none," and 5 signifying the highest rating, "a great deal."

How close did the Evaluation Summit sessions match what you expected?

In terms of matching expectations, the average rating for the Evaluation Summit sessions overall was 3.9 (nearly a lot), with 4 percent of respondents giving a rating of 2, 24 percent of respondents giving a rating of 3, 52 percent giving a rating of 4, and 20 percent giving a rating of 5. Four respondents indicated that they had no prior expectations and two respondents indicated that the sessions exceeded their expectations.

How much did you gain from the large group sharing of projects?

The average rating on the large group sharing of projects was 4.6 (nearly a great deal), with 2 percent of respondents giving a rating of 2, 31 percent of respondents giving a rating of 3, 41 percent giving a rating of 4, and 27 percent giving a rating of 5.

How much did you gain from the break-out sessions?

The average rating on the break-out sessions was 4.0 (a lot), with 2 percent of respondents giving a rating of 2, 20 percent of respondents giving a rating of 3, 47 percent giving a rating of 4, and 31 percent giving a rating of 5.

How much did you gain from other opportunities for networking?

The average rating on opportunities for networking was 4.0 (a lot), with 2 percent of respondents giving a rating of 1, 2 percent of respondents giving a rating of 2, 13 percent of respondents giving a rating of 3, 40 percent giving a rating of 4, and 38 percent giving a rating of 5.

How much did you gain from the Evaluation Summit overall?

The average rating on how much was gained from the conference overall was 4.2 (a lot), with 2 percent of respondents giving a rating of 2, 16 percent of respondents giving a rating of 3, 43 percent giving a rating of 4, and 39 percent giving a rating of 5.

Participants provided the following comments on what they gained from the Evaluation Summit overall:

> *I wrote down many good ideas and tips that I'll follow-up on.*
>
> *It was a great opportunity to share related work cross projects. I think that the conference will ultimately improve the quality of MSP evaluation.*
>
> *Trust and engagement and scholarship characterized the entire three days. Thank you.*

I received good feedback on my own work and made some connections with people doing related work who would like to build on the work I'm doing.

This was my first in this arena – it was well planned and structured as to relevant topics.

I would strongly suggest that you repeat this conference yearly. All evaluators with whom I spoke indicated the importance of this sharing of ideas, strategies, finding, and difficulties and the wealth of creative ideas to address the variety of difficulties.

The sessions sparked many emerging insights later during informal conversations.

How useful did you find the conference overall?

The average rating on the usefulness of the conference overall was 4.0 (a lot), with 2 percent of respondents giving a rating of 2, 17 percent of respondents giving a rating of 3, 48 percent giving a rating of 4, and 33 percent giving a rating of 5. Participants provided the following comments:

Too much being talked at, not enough time to interact.

There was a sub-theme of research methodology that was very interesting.

It was very good to hear about the many kinds of data collection and analyses that are going on in other projects.

The discussants were very helpful as timekeepers. However, their comments regarding the presentations were the least helpful I heard in the break-outs.

I would have liked there to be more time for small group interaction – conversations of attribution, challenges, tools, etc.

The timing of the conference was great! It pushed us to pull results together over the summer, to write them up and reflect as we move into our next year of work.

I found the willingness of the participants to share information and knowledge very refreshing – in contrast to the dozens of other institutional instruction conferences I have attended.

It was a nice balance of plenary and breakouts.

Great variety of sessions/formats kept things interesting and engaging.

I will be using and reviewing what I heard here for years! The personal connection for our team and for other teams are also a great outcome.

I heard of very little "research" in this conference that is anywhere close to being acceptable/competitive for publication in good, refereed science education and math education journals. What does that mean?

What are the most important and relevant topics that should be discussed at a future Evaluation Summit?

Participants provided the following responses, arranged according to topic.

Partnerships:

Partnerships, attribution between program interventions and outcomes.

Continue conversations about partnerships.
Defining partnerships.

I'd enjoy it if you'd continue to follow the threads of emergent evaluation and partnership issues. Keep inviting us to share our works in progress so we can learn from each other.

Characteristics of effective partnerships, bi-directional learning.

Conversation should continue, perhaps through a threaded discussion on MSPnet, related to partnerships, what is it and how to evaluate it.

Higher Education/Institutional issues:

I recommend a strand that deals with institutional issues (both K-12 and higher ed) including institutional change, policy changes, and K-1g alignment.

Impacts on IHE.

More information on Higher Education outcomes and effects of MSP.

Examining the culture of IHE versus schools, colleges of education, and arts and science, arts and science, and administration versus teachers.

Collecting Evidence and Results:

What is good enough in terms of collecting evidence and development of instruments.

Techniques for collecting reliable data and for standardizing data definitions. Also, methods for providing unique data identifiers – classrooms, schools, districts.

More emphasis on data and conclusions – less on description of projects.

Specific evaluation or research studies grounded in the literature with attribution of results grounded in the literature.

Has evaluation and the evaluation process had effects beyond the MSP boundaries.

Continue to share project results.

The effectiveness of different methods to generate the evidence you need.

Connecting the dots between projects.

What is evaluation that really informs classroom instruction? What is evaluation that informs us of parents' role?

More information on what RETAs are learning or what serves they have been called on to provide.

Attribution issues on how we understand the contribution of MSP at the project level, at the program level, and at the national level.

Greater presentation of data – empirical – both qualitative and quantitative to support conjectures/conclusions.

The opportunity to discuss work in progress and to discuss the ways in which people are trying to make sense of findings as they emerge.

Linking of program measures to student achievement. Identification of good program measures.

Identifying characteristics/questions of success. Identifying/explaining personal challenges in implementing evaluation program with resistance, lack of teacher data, etc. and sharing ideas on how to deal with problem .

Methods and Design:

Purposes, drawbacks, and benefits of logic modeling.

How to expand sharing across projects, selecting and evaluating comparison groups, and statistical analyses techniques (novel applications).

Methods and design as well as power and limitations of different statistical methods.

Continued sharing about research designs, instruments, and analysis to complement the work of the Utah State RETA and WCER.

Additional evaluation summits with particular focus on instrumentation for classroom observations and strategies for formatting data/evaluation reports with meaning to users.

Data analysis techniques for value added approach and for non-experimental designs. As people move forward in the completeness of data collection, these types of methods will be crucial for future presentations.

Challenges of evaluating an MSP.

Sessions that focus on methods to get best practices into classrooms before the end of the MSP.

Teacher Content Knowledge:

The measurement of teacher content knowledge should be an entire strand.

Ways of measuring teacher change – especially content knowledge and classroom practice.

Cultural Context and Race, Ethnicity, and Underserved Students:

More about understanding impact on underserved students.

More information on underrepresented minorities in STEM.

Responsiveness to the cultural context in the evaluation and program development (race/ethnicity, gender, disabilities). Intersection of dimensions of diversity – rather than focus on a single dimension (e.g., race/ethnicity)

Other comments:

More informal meetings to share ideas with other MSPs will be beneficial.

Panel discussions around specific challenges confronted where participants can share and discuss.

Opportunities for publication of research/project outcomes.

I think more of the same would be great!

Require papers again.

A few participants expressed concerns regarding the discussants' role:

Presenters of research spent far too much time presenting an overview of their project and not enough on a particular research effort. Far too often after the presentations and comments by the discussant, there was little or no time left for an open discussion by participants in the room.

The three-panel organization was nice for the audience but not for the presenter, but overall it worked well. The discussant's role appeared to be more of timekeeper than discussant and that is as it should be. Overall, I found the audience could devise talking points, without the discussant having to do so.

Several participants made additional comments regarding the overall conference organization:

I recommend setting aside time for participants to reflect on the implications of what they are learning in the context of their projects.

This was a very nice sharing of work in progress. People felt comfortable sharing things that hadn't worked out well, or talking about uncertainties. This created a real sense of a community with people working together to understand something.

Overall very well done, congratulations to conference planners.

Thanks so much for a wonderful opportunity to learn.

Thank you for being vigilant about starting on time and allocating time carefully to presentations. I also appreciated having healthy choices in our means/snacks! Half-hour breaks are important for networking. Thank you for your hard work! Every aspect was enjoyable.

Very well organized. Very high quality of participants.

The timing of the evaluation summit might have been premature since presentations involved little discussion of findings and more discussion of methods/background.

Thank you for this wonderful opportunity to become a part of the MSP community. I am not a senior on the project and the funding provided by the RETA allowed me to attend.

Outstanding opportunity for MSP network sharing.

Thank you for pushing the scholarship of our work – the request for papers and the sharing of them.

Very nicely done – facilities, food, conference schedule, papers, etc. Excellent conference.

Summary

The 2005 MSP Evaluation Summit provided a variety of opportunities for both formal and informal conversations to identify critical evaluation needs in the review of the Mathematics and Science Partnerships. One of the highest rated sessions was the one on partnerships by Gordon Kingsley as well as the pre-sessions by Michael Patton on evidence-based evaluation findings and by Rob Meyer on statistical multilevel models. A summary of the evaluation responses indicates that the participants valued the Summit and gained new information. Based on the participants' comments and recommendations, in a future meeting, we will most likely continue to request participants to submit and present papers, since participants responded favorably to this aspect. We will continue to encourage the submission of early research findings, but also research that has a strong literature base. We will review the discussants' role to allow time for discussants to be more analytic, but also to allow time for participants' questions.