March 17, 2003

M E M O R A N D U M

To: Participants of the Adding Value to the MSP Evaluations Conference

From: Norman Webb, Principal Investigator

Subject: Summary of the February 20-21, 2003 Adding Value Conference

The first conference of the Adding Value to the Mathematics and Science Partnership (MSP) Evaluations Project was held on February 20-21, 2003 at the Wisconsin Center for Education Research, University of Wisconsin-Madison. MSP evaluators and participants present were:

William Badders, Cleveland Municipal School District
Phyllis Balcerzak, Washington University, St. Louis
Eric Banilower, Horizon Research, Inc., University of Rochester MSP
Kathleen Bocian, University of California-Riverside
Wesley Boykin, Montgomery County Public Schools
William Clune, SCALE, University of Wisconsin-Madison
Cynthia Copolo, Duke University
Mary Ann Gaines, Texas A& M University-Commerce
Kathleen Haynie, Kathleen Haynie Consulting
Daniel Heck, Horizon Research, Inc., Indiana MSP
Michael Howard, Michael Howard and Associates
Frances Lawrenz, University of Minnesota
Michael Mikusa, Kent State University
Susan Millar, SCALE, University of Wisconsin-Madison
Daphne Minner, Educational Development Center, Newton, MA
Colleen Orsburn, Vermont Institutes Evaluation Center
Paulette Poncelet, Cleveland Municipal School District
Edys Quellmalz, SRI International
David Ramirez, Center for Language Minority Education and Research, Long Beach, CA
Stephanie Schneider, University of California, Irvine
Lynne Tingle, University of North Carolina-Greensboro

Attendees from the Wisconsin Center for Education Research Adding Value Evaluations project were Rob Meyer, Norman Webb, and Paula White. This memo summarizes progress made at the conference.

Introductions and Charge for the Conference

Norman Webb, Principal Investigator of the Adding Value to the MSP Evaluations Project, Wisconsin Center for Education Research, gave an introduction highlighting the three goals of the conference: 1) networking among MSP evaluators and RETAs, 2)

identifying critical evaluation needs, and 3) providing information on designing evaluations. The conference was held at the initial stages of the evaluation of the MSPs. The expectation for the conference was not that evaluators would have a well worked out evaluation plan. Instead, the hope for the meeting was that through interacting with other evaluators, evaluation needs would become clearer. Following the introduction, conference participants were given a list of fifteen evaluation activities and discussed the evaluation methods they were familiar with and those they were not.

Project Overview

Norman Webb identified the goals and activities of the Adding Value to the MSP Evaluations Project. The three primary goals are: 1) to increase knowledge of MSP evaluators about design, indicators, and conditions needed to successfully measure change in student learning over time, 2) to develop useful tools for evaluators to attribute outcomes to MSP activities, and 3) to apply different techniques for analyzing the relationship between student achievement and MSP project activities to evaluate the success of MSP projects. The three key activities of the project are:

1) Technical Assistance to MSP Evaluators
   - Provide technical assistance to approximately 15 MSP evaluators
   - Interact with evaluators on site about specific evaluation issues
   - Help improve technical quality of MSP evaluations

2) Develop and Provide Tools
   - Develop a Web-based tool for analyzing standards and assessment alignment
   - Design a value-added analysis process and tool

3) Produce Prototype of Large-scale Student Outcome Analysis
   - Value-added analysis on change in student achievement over time
   - DIF analysis of underlying constructs achieved by contrasting groups

Part I: Measuring Intervention

Based on experience in evaluating the statewide systemic initiatives, difficulties arose in collecting data to adequately document the interventions and measure the outcomes. This session was designed to formulate a discussion of how to describe and monitor the interventions. Conference participants met in three groups of seven to nine participants to discuss the following questions:

1. What are the key interventions of the MSP you are evaluating?
2. What key indicators will be used to measure the interventions?
3. What tools will be used to measure the interventions?
4. What assumptions do you have to make to measure the interventions?
5. What can you do to overcome difficulties in measuring the interventions?

Following the small group discussions, the three groups reported back to the larger group on a summary of their discussion. Each of the groups did not necessarily use a similar format to report back their discussions, nor do these notes represent final conclusions, but rather issues and questions identified by the participants regarding the evaluation of the MSPs.

Group 1, Part I: Measuring Intervention

Interventions:
- Students – kits implementation, student mentors, building community and family connections
- Teachers – institutes, link professors/teachers, discussions, pre-service experiences, mentoring programs, teacher leaders, kits training, capacity
- University professors/faculty – seminars, course development, institutional change/policy
- School leadership training – principals/teachers, curriculum leaders, capacity building
- Partnerships – building relationships, changing institutional patterns
- How is the curriculum being implemented – observation, teacher report, monitoring professional development

Indicators:
- Curriculum content analyses – degree of alignment of content to standards, level of use of kits, student achievement, kit specific scores, yearly scores, existing assessment
- Validating grades and tests
- How is the curriculum being implemented - discourse analysis, observation, teacher reporting

Tracking:
- Participation, quality, content, teacher knowledge/teacher confidence

Group 2, Part I: Measuring Intervention

Interventions:
- Professional development, in-service, pre-service, systemic

Measures/tools:
- Change in behavior, observational, learning log (web-based), self-video product, self-rationale standards based, self report, faculty protocol, on site monitoring, knowledge, instruction, exemplars, PDOP professional development tool developed by Horizon, national assessments written into evaluations to be used as alignment measure, teacher performance standards, learning effects, evaluating student work using rubrics, classroom based assessment, performance assessment

Outcomes:
- Narrow the gap between reality and perception, growth

Needs/Concerns:
- Random versus comprehensive, coaching versus evaluation – the idea of an outsider coming into the classroom, evaluation versus help and who has access to the evaluation – word "evaluation" raises red flags especially with unions, how much is enough? Sustainability, overlap of assessments, not such a good fit between standards and the state tests

Group 3, Part I: Measuring Intervention

Appalachian MSP:
- Development and implementation of content and pedagogy, idea of mentored implementation and efficacy of support mechanisms to put content learning into place, will use teacher partners and extend to supervisors

SCALE Project:
- Largely urban, LA Unified, Providence, Denver, Madison – building on pre-existing systemic work, key interventions: bring curriculum and resources through learning processes, already been doing this in 3 of these districts, providing immersion units focusing on cutting edge science, related to technology and sciences, improving quality of pre-service learning and increasing content knowledge, dealing with the gap for underrepresented minorities, three resources will be provided for the intervention: a system of indicators, case studies, and an ethnographic study of the process of building a partnership

Rochester:
- Curriculum implementation, cadre of school-based leaders and release teacher leaders, increase a shared vision of math, informal professional development in schools as a teacher to teacher strategy

South Texas:
- A process of vertical teaming from K-12 schools and university to improve math education K-16 and enhance the application of technology with regard to math, using the state standard scores to determine narrowing of gaps in math achievement, takes a bottom up approach to interventions

Montgomery Public Schools:
- Vertical integration, mentor teacher program, focus on math and science with more emphasis on science, enhancing instruction

Needs:
- Collaborative effort so everyone isn't reproducing similar efforts for the same tasks

Part II: Measuring Student Outcomes

This session was designed to discuss evaluation efforts to measure student outcomes including a variety of ways of thinking about student outcomes such as student achievement, student pursuit, and student activities. Conference participants met in three groups to discuss the following questions:

1. What are the expected student outcomes?

2. How will the student outcomes be measured?
3. What variables do you expect to detect to attribute changes in student outcomes
    to the interventions?
4. How will the changes in student outcomes be related to the interventions?
5. What is the timing of the outcome efforts?

Group 1, Part II: Measuring Student Outcomes

STUDENT OUTCOMES:	Measurement/Attribution
Rochester:	
Student achievement	State 4th, 8th
(gaps; gender/ethnic)	
Attitudes	
Course selection	Course patterns
Specific exams	Math A HS proficiency
Texas:	
Student	Exit level exams, state mandated exams
End of course	Final exams
Course selection	Course patterns
College prep	Follow post-high school career
Maryland:	
Achievement	State mandated
Attendance	All grade levels
Disciplinary	(gaps)
Montgomery:	
Close achievement between low	
schools/high schools	
Achievement	State mandated
High school assessment	High school science assessment
Biology	NCLB 10th grade
	# of students enrolled in honors, AP science
College grads/teaching profession	Track math, science grads enter
	Teaching, increase hires from
UC Riverside:	
Critical thinking	
Motivation	
Cleveland:	
Achievement	Changing state mandated tests, SAT-9
Course selection	Course patterns
SCALE:	
Challenging content	Use sample approach
Attitude (willingness, persistence)	
Students as mentors	How to measure depth of explanation
Dropout rate	Standardized
Pre-service teachers	Content and pedagogical knowledge
Form of activities in classroom	

Attribution:
With multiple external programs how can we attribute gains, differences to the MSP?
- Value-added analysis
- Causation mapping
- Isolation, comparisons
- Critical measures
- Track indicators and intervention over time
- Levels of intervention (concentration of activity)
  -schools as unit of analysis
  -measures of school climate/support
  -change in performance over time
- Generic definitions of teaching and student outcomes versus differentiated instruction

Tracking issues:
- How to reliably track school-based interventions?
- Reality of a control group, teacher volunteers vs. recruits?
- Teacher effect versus school effect

Controlled studies:
- Link professional development→teacher behavior→student outcomes
- Sampling process (random) – observations – student perception – achievement
- Teacher perceptions
- Numbers? (*n* needed?)

Artifacts:
- Student work, teacher logs (accuracy, usefulness), intermittent prompts – record behaviors
- How to measure quality of instruction
  -Differentiation – how to measure?
  -Expectations – how to measure?
- Observation instruments should: indicate grouping, track student engagement, change of teacher strategies, student response, qualitative notes

Control groups:

MSP	Control?	Random?
Baltimore	Yes	No/Maybe
Rochester	Maybe	Maybe
MCPS	Maybe (union, volunteer)	No
Stark County	Maybe	No
St. Louis	Yes	Maybe

Outcomes:
- Student performance - state assessments, content specific assessments, national standardized tests, student classroom processes
- Student access - A/P and advanced courses, gate-keeping (e.g. Algebra 1), elective courses
- Post secondary
- Teacher pre-service
- Majors
- Career choice

Tools:
- Change in performance level, correlating items with professional development experiences, content and process items on tests, classroom observations, journals, look at group differences, course availability, course support

Attributing change:
- Level of school-based intervention to level of student outcomes
- Causal model – statistical-based methods - how to show true success if your intervention is not the only program
- Matching schools on demographics, compare change in MSP school
- Identifying successes

Timing:
- Annual assessments (spring administration)
- Perhaps more than one year out? (summer institute impacts)
- Annual classroom visits
- Some to be developed, implementation one year away

Group 3, Part II: Measuring Student Outcomes

Student Outcomes:

What?	Level
Student achievement (state test)	
Math skills/knowledge	
-application/concept	Grade 3, 6
Pursuit/course taking	Grade 8-12
Achievement (math/science)	Grade 3-8
Achievement	
-math/science (mult/SA/Ext)	Grade 10-12
Interest (math/science)	
Enrollment in higher level courses	
(math/science	Grade 9-12
Achievement (math/science)	Grade 3-8
Participation	
Disposition to learn	

How and When?	**"Grain size"**
State test	Individual
Annually ("treatment and control")	
Transcript	Individual
State test (when in flux)	State, want to get individual
District tests	Individual
State proficiency test	
Survey/inventory	Individual
Transcript	Individual
State test (annual)	Individual
Project constructed test	
Enrollment	
Matriculation to higher education	District
% graduating	School/district
% tested	School/district
Attitude inventory	
Parent satisfaction	

Analysis Plan:
- Assess impact of specified professional development and curriculum implementation (of a new curriculum)
- Through change in student achievement
- Assess impact of variety of professional development experiences of teachers
- Assess impact of professional development of teachers
- Impact of teacher professional development experiences
- Including disaggregation

Part III: Assessing Causation

Attribution was extremely difficult to measure with the systemic initiatives. This session was designed to discuss the importance of setting up the evaluation in the beginning of the intervention with tools to determine the impact of the interventions. Rob Meyer of the Adding Value to the MSPs Evaluations Project, gave a presentation on value-added evaluation strategies. Following Meyer's presentation, conference participants met in three groups to discuss the following questions:

1. What kind of data suggests different evaluation approaches?
2. What kind of data will you need?
3. What data do you have for relating implementation to student learning? What additional possibilities do you have?
4. What evaluation strategies could you use if you collected this additional data?

Group 1, Part III: Assessing Causation

- What kind of data? Control groups, past performance data, pre-service component

- Differences between evaluation and research, trying to find out what evaluations can be done in a research frame, a school impacted in year is not necessarily affected the same way the next year, treatment verification

Grade levels – outcome data:

Baltimore County:
- All grades, 3 years before implementation, CTBS test, possibility of designing some authentic measures, the goals for the students is a 10 percent increase across the board, will be able to link the data across grades since it will be longitudinal data, a series of math and science exemplars, a performance assessment that goes along with that and use them in the classroom, high quality assessment is expensive, difficult to balance no child left behind and high quality assessment

North Carolina:
- Pockets of folks who have used performance assessments for science, may be other pockets, if CTBS data isn't enough, what else will you put in? If you're going to report on multiple levels, what are the hybrid models so you can do some of this high quality assessment but not break the bank?

Rochester:
- 4th grade, 8th grade, Math A exam, plus all the Regents exams but only a small number of teachers teaching those courses

Stark County:
- 6th grade test data plus 9th grade achievement data which will be replaced with a 10th grade test, students can be tracked, but not sure how well correlated 6th grade test will be with 10th grade test, it would be nice for us to get some value-added data by doing a pre-post test with performance assessment by grade level, but difficult to manage with 15 districts over 3 years

St. Louis:
- State science assessments at grades 3, 7 and 10, math at 4, 8, and 11, we have district data from 1996, we're trying to get the data discs so we can manipulate it into a data base to track students, not sure if we can get this together before June, 2003 before MSP treatment starts

Maryland County:
- Program focuses on high schools, track students across grades and teachers, will use AP exams, enrollment in the honors and AP courses, will use MSPAP as well, some of the courses are grade level specific, tracking the students is even more important here since not necessarily taking the courses in traditional grades

Group 2, Part III: Assessing Causation

- 2 phases, one starting their treatment this summer, another next summer (so have control group), district has own essential standards based on criterion-referenced tests administered in fall, winter, and spring, biggest problem is number of teachers receiving treatment since spread out among schools, will do work on math content, problems in getting at critical thinking for students
- With 55 districts, the *n* isn't a problem, but evaluation team was brought in late in the implementation process, some random assignment possibilities, Inverness is

evaluating 3 projects and can do a subset of participating districts, will be a real challenge to tie intervention to outcomes
- Looking at professional development in our sites and how there are levels of involvement and how to track these

Group 3, Part III: Assessing Causation
- Group 3 took a case design analysis approach to get as much detail as possible on one MSP. The Cleveland MSP volunteered to be analyzed.
- Cleveland MSP, teacher-focused program, will initiate a master's program. Teachers are to increase their knowledge by attending sessions at a university plus a mentor program.

Considerations for evaluation include:
1. Causal modeling requires planning, looking at latent variables (up front)
2. Put resources into design to make it more testable (building on research)
3. Simpler is better, break model into parts, need to know how pieces fit together
4. Randomization at a key point can be a key to measuring effects
5. Identify the fatal flaws to see if these can eliminated by design (randomization)
6. Get to know the data system and who has access to the data building, relationship to district's data system, keep people informed of why and what
7. Need proximal and distal outcomes
8. Cost benefit analysis of evaluation analysis
9. Case based methodology works and is helpful – spending significant amount of time talking about one project helps to apply to another

Reporting Out: Identification of Evaluation Needs

Norman Webb asked participants how they would want to spend the two days of the next Adding Value Conference in September 2003. The following items were raised and in some cases, participants gave a show of hands as to the number interested in covering the items.

- Baseline data of student assessment – 6 people interested
- Measuring quality professional development – 6 people interested
- Focusing on formative feedback to projects – 2 people interested
- Issues of equity and diversity/interventions and outcomes – 6 people interested
- Design experiments

Last Comments

Bill Clune: evaluators will need assistance faster than next September, an information solicitation through the internet could be helpful, a threaded Web discussion, a list serve, would require someone to report on evaluation updates, make it specific

Michael Mikusa: a bulletin board is good, can read everything you want that's been posted on the topic

Frances Lawrenz: a telephone conference call, with a scheduled time of 2 hours, pre-read materials and specific questions, you could get 8 experts around the table at the same time, focused, very effective, it doesn't work well to do something that's vague, targeted questions, video links work well

Colleen Orsburn: for video conference, meet together then go away for a few hours and work on issues and then regroup

Kathleen Haynie: need a balance between topic driven and something more responsive to needs of individual projects, helpful to get responses to specific needs

Stephanie Schneider: later down the line we'll address the data analysis, not just the design, also helpful to use methods based case studies with a case analysis based on use of that method as a learning tool – that would be a great half day session next September -- a case analysis

Wesley Boykin: I'd like to see how a method is attached to a specific goal, most of the goals of my project aren't related to student outcomes, they're related to changes in faculty teaching

Kathleen Bocian: we've been lucky to have some district people participate, it would be nice to have people even from here locally so part of the discussion involves people from the schools – so can raise school level issues, their involvement would be helpful

Norman Webb: including PIs is helpful and on the NSF web site, a summary of all the projects and their goals can be used to look at the context of certain methodology

Susan Millar: in the first hour, Norman Webb made us do an activity, how to get evaluators talking, important to give each project a chance to be the focus, a round robin project critique

Michael Mikusa: meet in small groups so more people can have their own project discussed, if we hash out the details here but then don't hash out the problems with the folks back at the MSP, will the districts agree to do this?

Daphnie Minner: provide some guidance for predictable roadblocks, issues of getting some test scores, some predictable things we can alert folks to

Resources

http://education.umn.edu/CAREI/cetp/ - go to instruments – includes faculty, teacher, and principal surveys for higher education and K-12 education, classroom observation forms for K-16, and rubrics for student learning
http://education.umn.edu/CAREI/Programs/SSC/default.html - go to instruments - includes assessment instruments and teacher surveys

http://education.umn.edu/CAREI/SI/default.html - go to instruments - includes teacher, principal, and student surveys about their school environments for middle school and high school math and science
http://oerl.sri.com - Online Evaluation Resource Library
http://palm.sri.com - Online Interactive Performance Assessment Resources
Evaluating Schools, Programs, and Policies: Selected Bibliography (distributed at conference)

Please mark September 18-19, 2003 for the next Adding Value Conference.