

Evaluating the MSP Program: Looking at Return on Investment

2013 Math and Science Partnership Learning Network Conference
Implementation: From Vision to Impact
February 11-12, 2013 • Washington, DC



A FEATURED TOPIC SESSION

Presenter:

Mark St. John
Founder and President
Inverness Research Inc.

Contents:

Introduction	3
Some General Thoughts About Evaluation.....	4
Evaluating Investments in Educational Improvement.....	10
The Improvement Infrastructure	13
Group Discussion.....	20



2067 Massachusetts Avenue
Cambridge Massachusetts 02140
617-547-0430

TERC



MSPnet

An electronic learning community
for all MSP projects.
www.mspnet.org



This documentation is part of the **MSPnet** project, funded by the National Science Foundation and designed and facilitated by **TERC Inc.**, a not-for-profit education research and development organization based in Cambridge, Massachusetts.

Participant comments have been paraphrased; they are not exact quotes. The contents of this document do not necessarily reflect the views of TERC, the National Science Foundation, or the organizations of any participants.



MSPnet

An electronic learning community
for all MSP projects.

www.mspnet.org

Principal Investigators
Joni_Falk@terc.edu
Brian_Drayton@terc.edu

Conference Documentation
Catherine McEver

About This Summary

This document focuses on one of the Featured Topic Sessions at the 2013 Math and Science Partnership Learning Network Conference (LNC). The intent is to provide a summary of presentation highlights as well participant discussion.

Abstracts for these presentations may be found at the 2013 LNC site on MSPnet (see URL at right). A video of this session as well as the original PowerPoint slide presentations are also available on MSPnet.

The LNC Online:

http://hub.mspnet.org/index.cfm/msp_conf_2013

Readers interested in pursuing any of the plenary session presentations from the 2013 LNC are encouraged to access MSPnet to find a document summarizing these sessions, full video recordings of the presentations, the original PowerPoint presentations, and detailed speaker biographies. All abstracts submitted for breakout sessions during the conference are also available on MSPnet.

Introduction

Mark St. John notes that Inverness Research has been thinking about evaluation and how to evaluate NSF projects for decades, and this session provides the opportunity to think out loud, provoke reactions, and evoke some shared thinking.

The Poster Session Challenge

Imagine Congressional members come to the LNC poster session. They want to know “what their money is buying.”

What would they learn from the posters?

How well do the MSP projects conceptualize and communicate “return on investment” (ROI)?

He begins with a challenge, recalling an earlier Learning Network Conference in which Cora Marrett spoke and made a plea to the NSF community, asking for help regarding how to talk to Congress about MSP and what it has accomplished. At that session, St. John toured the project posters and asked himself, “What if I were a congressperson and had to make a deduction from going through these posters about what it is we are buying with the money for this federal initiative? I would have been mystified,” he observes.

It raised a question about the ability and wisdom of projects and evaluators being able to document and summarize the return on investments of MSP projects. Projects are very

sophisticated in talking about things like pedagogical content knowledge, teacher subject matter knowledge, etc., he acknowledges, but the return on the federal investment is not always clear. That is a focal point for this talk.

For the homework challenge, St. John asks participants to think about how well we can explain, grounded with evidence, the return on federal dollars of the MSP. What is the role of MSP projects in the overall educational reform landscape?

In talking about Inverness Research and explaining what he does St. James says, “We study the investments that are made in the improvement of education.” Each of those words is loaded, he notes. The question of what an investment is and what it means to invest in the improvement of education and to study that effort are focal points of this presentation.

Inverness Research has a history with Math and Science Partnerships as well as with many other



Inverness Research:

We study investments made in the improvement of education.

Inverness and the MSPs

- Rapid City (PRIME) evaluators for ten years
- Appalachian Math Science Partnership (AMSP) evaluators for twelve years
- Maine (PSP) evaluators for three years
- MSP Collaboratory documentors for one year

Summary:

Are the Mathematics and Science Partnerships an expenditure or an investment of federal dollars? Seen as an expenditure, the MSP grants buy services and short term outcomes. In this view MSP funds are used to enhance student achievement, in particular. An investment differs from an expenditure in that it produces capital—enduring assets that can be used for future improvement efforts. Too often we believe MSP projects do not see themselves as long-term investments. And evaluators do not carefully document the different forms of educational capital the MSP projects create.

This session will explore the need for documenting and communicating the total return on investment that accrues from the MSP investment. Discussions will focus on the ways in which evaluators can broaden their lens and capture long-term capital creation as well as short-term activities and contributions.

More information regarding Mark St. John and the work of Inverness Research Inc. may be found at:
www.inverness-research.org

A Play in Three Acts

- I. Some General Thoughts about Evaluation and Evaluating MSPs
- II. Evaluating Investments in Educational Improvement
- III. The Improvement Infrastructure

NSF initiatives. They worked with the Appalachian MSP for over a decade and with the Rapid City (PRIME) MSP for ten years, and just wrote a report called “Ten Years of PRIME.” They are now the evaluators for the Maine Physical Science Partnership (PSP) and are also working on the new MSP Collaboratory.

This working session consists of a play in three acts, St. John explains, beginning with some thoughts about evaluation from the Inverness Research perspective.

Some General Thoughts About Evaluation

St. John begins with thoughts about evaluation that he characterizes as “a little overstated and provocative.” First, he posits, evaluators are

not in the business of evaluating education or studying education and how it is done. They are in the business of studying incremental investments, external investments that are aimed at improving education, and improvement is a different process than education, he notes. “We need to study how you improve things and the strategies and outcomes that are appropriate,” St. John explains. “We need to keep our lens on the fact that we are investing in and studying efforts to improve education.”

The second point is that Inverness Research and others evaluating MSP projects are independent evaluators. The real value in being an independent evaluator comes from the fact that you are neither of the foundation, nor of the project, St. John states, and your allegiance is to the investment. What does this mean? When a project hires you they want you to do certain things. You want to listen to them and work with them, but as an independent evaluator you need to assess the value of this investment independently. Your allegiance is not to the project. At the same time, the foundation may have its own ideas about what you should be doing but again, you need to make it clear that the independent evaluator’s role is to be a third point on a three-point triangle: the project, the funding source, and the evaluator. “Your ultimate allegiance has got to be to stewarding, protecting, and understanding the investment that is being made,” St. John emphasizes.

Another assertion is that evaluation is fundamentally about understanding what is

Evaluating Improvement Efforts:

- *The process of improving education is a different process than doing education ... requiring a different set of skills, foci, and responsibilities.*
- *The allegiance of the independent evaluator is to the investment, not the project or the Foundation.*
- *Evaluation is fundamentally about understanding and describing what is actually happening.*
- *Evaluation is NOT about assessing the degree to which stated goals have or have not been met.*
- *Evaluation should generate insights as much or more than “proof.”*
- *It is often more important to help projects think insightfully about their work and the situations they face than to address their initial worries of garnering proof of effectiveness.*
- *Evaluation also involves coming to understand the value of something... where value is ultimately multi-dimensional contextualized and relative.*
- *Evaluation is evidence based but is not equivalent to measurement or the accumulation of data.*
- *The thinking, judgment and perspectives of the evaluator can be as important as the data collected.*

happening. That may sound easy or trivial, but it is not, St. John observes. It can be very difficult at times to know what is actually happening. Projects often have a logic model and theory of action, and we know what is supposed to be happening, but what actually happens may not closely resemble what is supposed to happen. “Evaluation is about uncovering, documenting, and portraying what is actually happening, not only in terms of outcomes but in terms of processes, in terms of thinking, and in terms of interactions,” he states.

Evaluation is *not* about assessing the degree to which stated goals have or have not been met. St. John cites one of his mentors, Michael Scriven, who referred to something called the “goal for the evaluation,” and who said, “It is important not only not to study the goals, it is important not to know the goals.” Think of Consumer Reports, St. John suggests. When they are rating a car they don’t go to General Motors and ask what their goal was for the car and then evaluate whether or not General Motors met its goals. They take the criteria relevant and important to consumers and apply those to the cars. In the same way, St. John points out, it may be good to know a project’s goals, but that is not the evaluator’s job. The evaluator’s job is to assess value and use criteria that are relevant to the consumer.

As an evaluator you don’t want to be swayed by the project’s goal because what actually happens might be very different. St. John uses the Rural Systemic Initiative as an example.

They started with the goal of building catalyst schools and had a theory for why those were important. It turned out the key thing wasn’t catalyst schools, it was teacher partners and a network of leading teachers that were the essential element. Had Inverness Research set up ahead of time to study catalyst schools, they would have missed the entire initiative.

St. John observes that evaluation has been hijacked by the “proof camp,” those who believe the purpose of evaluation is to provide proof or to disprove, to do a test regarding whether something works or not, yes or no. “That is not the purpose of evaluation,” St. John says. “Evaluation should generate understanding of what is actually happening. It is as much about creating insights into the processes of the theory of action, into what is actually happening, into the design and the implementation, as it is about providing proof. I believe the real value of evaluators comes from not only gathering evidence about ‘effectiveness,’ but also helping people think about the project more powerfully, more succinctly, more accurately.”

When you get insightful, you see a project differently and have a different way of conceptualizing it, St. John explains. That insight helps people adjust their mental models and their thinking about design and implementation. Ultimately, evaluation is about helping people understand the value of something. There are many dimensions to value. Think about what you value in a car, for example. Evaluation provides information and insight

Mark St. John



impact
n ['ɪmpækt]

1. the act of one body, object, etc., striking another; collision
2. the force with which one thing hits another or with which two objects collide
3. the impression made by an idea, cultural movement, social group, etc. the impact of the Renaissance on Medieval Europe

Mark St. John



about a car, a project, or a process so that you can value it more accurately.

Evaluation is *not* purely data gathering and measurement, St. John stresses. The evaluator's job is not about producing data or measurement, it is about providing insight and understanding. That understanding should be evidence-based, with arguments behind it, but the job goes beyond gathering data.

The word "impact" has to do with an overarching metaphor, St. John observes, and when you have an overarching metaphor it often shapes your thinking and the terms you use. This LNC's theme is "From Vision to Impact." "Impact" is a funny word, he notes, having to do with collision, and a collision is short-term and violent. "Is that what we want for our projects, short-term, hammer-like impact, where things change? I don't think that is the right metaphor for the investments that NSF is making," he states. In looking for impacts we are looking for short-term jumps or changes, which comes under the metaphor of fixing a problem: If we have a project that has a problem, we fix the problem. That is not the way things work, St. John points out.

Finally, when you hire an evaluator you are hiring their thinking, their judgment, and their perspectives. An evaluator is like a connoisseur who knows the theater and can help illuminate and disclose that theater through the art of criticism. This means, he notes, that evaluators have to be good, but it also means you probably shouldn't evaluate in a domain you don't know.

St. John moves on from talking about the overall art of the evaluator role to NSF investments in education.

INVERNESS
RESEARCH

*Our approach to
evaluating NSF
investments in
education.*

When you evaluate an NSF project there are four possible functions you might undertake. One is foundational, that of documenting and portraying what the project is and how it works. With some very complex projects, he observes, that is almost all you can do—translate for others what the project is and how it works.

Four Functions of Evaluation

FUNCTION	AUDIENCE	PURPOSE
<i>Document and Portray</i>	<i>Internal and External Audiences</i>	<i>To help both insiders and outsiders better understand nature and purpose of project</i>
<i>Formative Feedback</i>	<i>Project leaders and staff</i>	<i>To help the project learn about its design and impact and thereby revise its design and strategies</i>
<i>Summative Assessment</i>	<i>Funders</i>	<i>To help funders assess the ways in which and the extent to which the project is creating value... and to assess the return on their investment</i>
<i>Research</i>	<i>The "Field"</i>	<i>To generate knowledge and insights about the improvement of education</i>

A second function is formative feedback. Here, the client is the project itself and the evaluator's job is not proof, it is to improve by feeding improvement information back to the project. "I would argue that for NSF projects, we should

put a lot of our efforts into helping the projects along the way,” St. John states. “Projects are hard and MSP projects are very hard, with complex partnerships. Helping a project through your thinking, your judgment, your perceptions, and through evidence, is an important role for the evaluator to play.”

A third role is summative evaluation, a classic role in which the client is the foundation or the foundation’s audience, which may be Congress, etc. They want to know whether it worked, but the way in which they ask that question is usually not productive, St. John points out. They ask, “Did we meet our goals? Was it effective? Was there proof of student achievement?” The latter, he notes, is the holy grail these days. Maybe we can prove it, maybe we can’t, but proof is very difficult in this field and it is a difficult field in which to claim causal attribution. Few would go in front of a jury and say, “Yes, we have causal attribution,” he observes.

St. John proposes two arguments regarding summative evaluation. The first is that these limited definitions of summative are counter-productive. The other is that the real question for summative is not whether you reached your goals or improved student achievement, it is the return on investment. Conceptualizing the contributions of a project in multiple ways, particularly capacity building, is very important.

The fourth role for evaluation is research. This is when you decide to treat the project as a case of something (e.g., the Appalachian MSP as a case of building a rural partnership). The

project is used as a case to illuminate some more general principles or findings, and you are contributing to the knowledge output of the project.

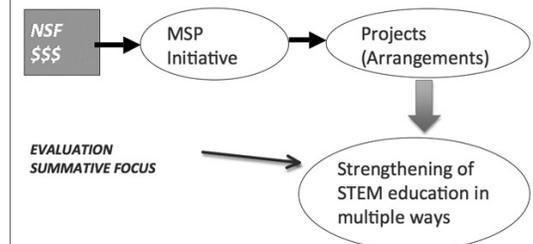
In the LNC plenary session, Joan Ferrini-Mundy commented on NSF’s original mission regarding human capital. St. John refers to NSF’s charter statement, noting that in the original days of NSF it was fifty-fifty between education and research. Now education is seven or eight percent. The important point is that the charter asks NSF to create things like the MSP initiative, and to fund creative projects that “make arrangements” (e.g., partnerships, networks, learning communities) to strengthen STEM education. “This is a very different metaphor from impact,” St. John stresses. “It doesn’t say ‘impact’ science education, it says ‘strengthen.’ As evaluators, it is our job to ask what that means. Are we able to evaluate the strengthening of STEM education?”

MSP has stated goals (see sidebar page 8), including things like challenging curriculum, increasing the number of math and science teachers, supporting the mathematicians and scientists involved, studying how students and teachers learn, and so on. “This is a fine set of goals for design and things to emphasize,” says St. John, “but it isn’t a sufficient template for doing an evaluation. You want to know what is the total return of the MSPs. This is an incomplete set at best. It is good to know these, but you don’t want to start your evaluation from there.”

The NSF Charter: Educational Mission of NSF

*The Foundation is authorized and directed to initiate and support basic scientific research and programs to **strengthen scientific research potential and science education programs at all levels** in the mathematical, physical, medical, biological, engineering, social, and other sciences by **making... arrangements** to support such scientific and educational activities.*

Evaluating NSF Funding



MSP Stated Goals

- an incomplete set of things to look for
- Enhance schools' capacity to provide **challenging curricula** for all students and encourage more students to succeed in advanced courses in mathematics and the sciences;
- Increase the **number, quality and diversity of mathematics and science teachers**, especially in underserved areas;
- Engage and **support scientists, mathematicians, and engineers** at local universities and local industries to work with K-12 educators and students;
- Contribute to a greater understanding of **how students effectively learn** mathematics and science and how **teacher preparation** and professional development can be improved; and
- Promote **institutional and organizational change** in education systems – from kindergarten through graduate school – to sustain partnerships' promising practices and policies.

St. John summarizes the points made during the first stage of this presentation.

Summary: An Assertion

- Evaluation of MSPs (externally funded projects) should document and communicate to both internal and external audiences:
 1. what is actually happening;
 2. the multiple contributions of the project which collectively comprise the total return on investment.
- This is different (much broader) than assessing distal contributions to target audiences.
- Evaluators should work with the project and the funder to increase the total ROI.

In light of this perspective on evaluation, St. John asks session participants to discuss among themselves the following question.

Question:

To what extent does your evaluation approach reflect or differ from this description?

Feedback from Discussion Groups**Government Performance and Results Act**

- We talked about the Government Performance and Results Act and people using a logic model to identify the outcomes they expect. We didn't feel that got to the values part as you described it. We are to the point of knowing what is happening in the project and helping people make decisions and whether they are achieving the outcomes they wanted, but we wouldn't interpret that to mean return on the value of the investment. • Participant
- The work is pushing us towards measuring things. It is why I would argue that as evaluators we have to have our own compass or North Star to guide us and push back. • Mark St. John

Logic Maps: Shooting Yourself in the Foot

- We discussed whether, independent of the evaluator, you should or should not know

Participant discussions



your goals. NSF, by its very nature, requires a well-developed logic model to get past the first reading. In order to build a logic model, the only way it is going to make sense is if the evaluator is part of that original logic model. In our project, for example, you have five project years, with linkages across each of those five years and implications for the data that become available. Within each year, research projects are supposed to be developed. The evaluator can play a key role in helping to define what those are. • Participant

• I agree. A logic model is a very important map, and the evaluator should be part of the mapping team, understand the map, and know what the goals are. What I would disagree with is when you say the map is real, or the map is fixed. Our job is to constantly update the map and find things we didn't know were on the map. • Mark St. John

• But NSF wants to see the map across five years. • Participant

• This is my point. You've got to stop the foundation from shooting itself in the foot. They are going to ask for things that are not in their interest. As an evaluator, I have seen foundations be their own worst enemy, and an evaluator has to have an allegiance to the investment, not to the foundation. If the foundation is asking for things that are counter-productive, you need to tell them it is counter-productive. • Mark St. John

Political Push for Measurement

• You can see why we got ourselves into this position of being pushed politically into having to measure everything because we went so far down a road of measuring nothing. In my state, when we tried to look at the effects of the AMSP, we obviously wanted to look at all the interventions the teachers are required to go through, not just their participation in AMSP. We found that states are spending millions of dollars a year and don't know where teachers are getting their professional development; all they know is that it is a requirement and the state is paying for it. Now the state has awakened and they want evidence that what they are spending their money on is actually generating something of value. The political pressures are real. • Participant

• They are very real. As an evaluator I am not against measurement and I am certainly for evidence. What I am against is measurement that is not evidence. We've become enamoured with measurement and data to the point where it doesn't make sense. I'm just for doing things that make sense, having arguments that make sense, coming back to my basic premise that we want to have an understanding of what is happening. Measurement can be very helpful in understanding what is happening. I am against measurement as a kind of end in itself, as opposed to a means to gain greater understanding. • Mark St. John

Loose Causality

• We were talking about the loose causality between what the project does and the claims they make about changes. It is not the way science is typically done, where there is lots of replication and you get to see if you still get the same changes if it is repeated with a different group. Norbert Wiener, founder of cybernetics at MIT, was home schooled by his father, and he was a genius. His little brother, who was home schooled even better according to his father, didn't amount to anything. • Participant

• The whole attribution-replication model is one I have a great deal of problems with, which is why I'm suggesting a slightly different approach and nuance. Context trumps everything. • Mark St. John

Participant discussions



Return on Investment as an Outcome

- Would you be adding return on investment as an outcome of a project? • Participant
- It is *the* outcome. • Mark St. John

What is “Capital”?

- The accumulated wealth of an individual, company or community, used as a fund for carrying on fresh production
- Wealth in any form used to help in producing more wealth
- Accumulated goods devoted to the production of other goods
- Assets that yield income and other useful outputs over long periods of time.
- (Expenditures produce products and services; Investments yield capital and generate working assets)

Evaluating Investments in Educational Improvement

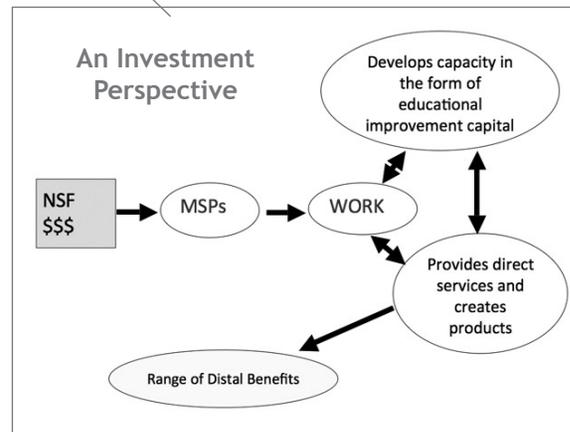
Launching into the second act of the presentation, St. John begins by clarifying terms. He starts by noting the difference between an expenditure and an investment.

Expenditure vs. Investment

- Expenditures are outlays for products and services.
- Investments create capital that can be used in the future production of goods and services.

“Expenditures are a one-time deal. If NSF is truly making investments, it has got to be doing more than making expenditures,” St. John asserts. He then proceeds to define the term “capital.”

NSF is funding MSPs, and those MSPs are doing a lot of work. This work does two things,



and the two things are symbiotic, St. John explains. One is develop capacity in the form of educational improvement capital. In the AMSP, for example, the project built capacity in the administrators and teacher leadership. At the same time, MSPs provide services, doing professional development, disseminating curriculum, building tools. These two things are symbiotic in the sense that the more capacity you have, the more work you can do, and the more work you do (if you do it correctly), the more capacity you will build. “We should be thinking of our projects as this dual, interactive process of building capacity while we are doing work, and doing work while we are building capacity,” St. John states. “All of that leads to distal benefits: raised student achievement, better instruction, etc. Evaluation is about understanding this diagram, it is not simply about measuring the distal benefits.”

St. John elaborates on what he means by educational improvement capital, pointing to different forms of improvement capital. He be-

Improvement Capital
Developed Through MSP Funding

Some Possible Forms of Improvement Capital

- Human Capital (People)
- Social Capital (Personal and institutional connections)
- Knowledge Capital (Ideas, knowledge, sharing and dissemination)
- Organizational Capital
- Financial and Political Capital

gins with a focus on human capital and various forms of human capital.

Human Capital (People)

Human Capital refers to people who have the expertise, propensity and position to assume leadership roles and contribute to the improvement of education.

Examples of Human Capital

- Science and mathematics faculty
- Department of Education faculty
- Teacher leaders
- Administrator champions
- Researchers
- Informal science educators
-

St. John offers an example of measuring human capital development. In studying the Nanoscale Informal Science Education Network (NISE Net), the evaluators asked a question early on: How many people in your institution have the ability, propensity, skill, and position, to do exhibits and programs in nanotechnology in your museum? The answer was 20, St. John reports. Four years later they asked the same question and the answer was 80, which means four times as many people in the informal world were positioned to work on nanotechnology. That was a measure of capacity building, which wasn't a stated NISE Net goal, St. John notes, but the evaluators highlighted this as a long-term contribution of NISE Net.

Next is knowledge capital, which is what the research function of NSF is all about, St. John observes. The MSP projects are expected to generate knowledge. Knowledge capital can take a variety of forms.

Knowledge Capital

The asset that is represented by the capacity to generate, share and more broadly disseminate knowledge.

Knowledge capital includes informal sharing of "craft knowledge," project knowledge about design principles, and evaluation and research design knowledge. "Are we capturing this as evaluators?" St. John asks. "If we are not capturing this, I would argue we are missing a big piece of the return on investment."

There is also social capital, and St. John points out that in recent years, social network analysis and diagrams have become important. Why? Because of the belief that there is an asset in the fabric of social relationships. In the MSP you can look within projects and across projects, and in many cases social capital is a project outcome.

Social Capital

Social capital refers to the asset of connectivity and to the capacity that comes from the connections and relationships that are developed through the project.

Organizational capital strengthens STEM

Examples of Knowledge Capital

- Publications
- Curriculum materials
- Online courses and tools
- Website(s) and media
- Identification and sharing of "best practices"
- Informal "Craft knowledge"
- Project -- Design principles and knowledge
- Evaluation and research design knowledge

Examples of Social Capital

- The personal relationships formed
- The relationships created in the form of partnerships and collaborations
- Development of local and regional communities and networks
- Connections built across the field through websites and on-line communities
- Relationships developed with associations and funding agencies (e.g. NOAA, NSF, NSTA, AAAS, NCTM...)

Organizational Capital

Organizational Capital refers to those institutional capacities and organizational structures that support STEM education and the improvement of STEM education.

Examples of Organizational Capital

- New organizational structures which could include new centers, alliances, networks
- New Institutions
- New Centers
- New Departments, Positions
- New programs and courses

The Overlap and Mutualism of Different Forms of Capital



education by strengthening the organizations involved and by creating new organizations or new partnerships and networks of organizations. For example, the Maine PSP project has four new tenured faculty positions at the University of Maine devoted to STEM, which represent new organizational capital, St. John reports.

Then there is political and financial capital. “This means you have raised the capacity of the project or region to have political leverage and to raise finances,” St. John explains, noting that NSF grants should better position projects and entities or networks created by projects to fashion a supportive policy environment to generate better policies and raise necessary funds to implement those policies.

Political and Financial Capital

Political and Financial Capital refers to those assets that help create a supportive context for STEM Education.

Examples of Political and Financial Capital

- Supportive policymakers and policies
- Supportive policy environment
- Multiple funding sources
 - National, state and local levels
- Capacity for raising funds
- Capacity to shape policy and finances

These forms of capital are not distinct, St. John notes, they intersect and overlap and mutually reinforce each other.

St. John then asks participants to discuss among themselves their own projects and consider whether they are attending to and capturing these forms of capital as part of the total return on investment.

Question:

What forms of capital are being developed by your MSP?

Feedback from Discussion Groups

Attribution and Specificity of ROI

- One thing that was brought up was attribution. As a more quantitative person, I am thinking about your term, “return on investment,” more literally. How do you attribute some of these changes in human capital? How do you define return on investment, or do you not mean it literally? • Participant
- I don’t mean it to the point where we can put a dollar figure on it (e.g., for every dollar spent we estimate a \$3.80 return). I am at the first order of beginning baby steps here. I would like to say that when NSF invests in the MSP they are not only buying professional development, they are not only buying curriculum implementation. After five years we see some increases in the capacities that I’ve talked about. Can I attribute them completely? No, that is why I use the word “contribution.” Another way to think about this may be “probable cause” versus “beyond reasonable

doubt.” There is probable cause that we have made some contributions to leadership. Many in the NISE Network participated, learned, built nano exhibits, and now they say they’re ready to do this kind of work. That is probable cause in my mind. Could they have learned it somewhere else? Yes, but there is probable cause. • Mark St. John

Helping NSF Appreciate Its ROI

- A simpler way of saying this may be that NSF makes these investments and then asks for evaluation that looks at these downstream distal effects, and they are missing a lot of what they are getting. They are putting money in and getting a lot of return that we are simply missing. And if they say, “We are here to increase student achievement,” I think that’s the wrong argument, the wrong role for NSF.

Evaluators have to help create the right frame, the right set of expectations, the logic theory, and the outcomes that actually match the reality. If we don’t put in the creation of capital, the creation of working assets, we are missing a lot of what they are doing. • Mark St. John

The Economist Approach

- In the most recent edition of the *Journal of Education Finance* you will see how to do that very narrowly defined student achievement and the dollar return on it from the MSP.
 - Participant

- Economists are fearless. • Mark St. John

Getting Credit for Important Outcomes

- I think all of us are probably doing some of this, we are just not capturing it in our evaluations. We see partnerships being formed, we see capacity building among our partner organizations, but we have been so focused on the goals of the project that we have overlooked these really important pieces. I think it’s good to bring us back to this subject. • Participant
- It is real work with real benefits, and we ought to get credit for it. It isn’t just an aside or a means to an end, it is an important outcome. • Mark St. John

The Improvement Infrastructure

Building on the thoughts presented, in act three of his talk St. John introduces the concept of the improvement infrastructure and observes that it is a useful concept for projects, for evaluators, and for NSF. In the LNC plenary session, he recalls, Jim Spillane talked about the importance of infrastructure. By investing in infrastructure such as a power plant or a freeway system, you get an asset that does future work.

When you invest in an infrastructure you are investing in something that addresses an important need. An infrastructure investment empowers a range of other activity. For



Participant discussions

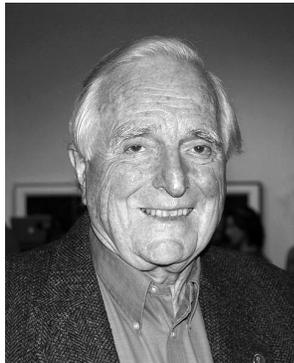
Infrastructure Investments



Investments in Infrastructure

- Address an important need
- Empower a wide array of context-specific local activities
- Bridge critical disjunctures
- Provide for equitable access
- Assure consistently high-quality services
- Allow for cumulative growth and development of capacity (scale up or down as needed)
- Are cost-efficient, with multiple sources of funding
- Provides a vehicle for future efficient investments

Doug
Engelbart



Wikipedia open source photo

Every organization has a “capability infrastructure”—what we use to do our jobs.
(Jim Spillane talked about this in LNC plenary)

Organizations also need an “improvement infrastructure”—what we use to get better at getting better.

example, when you invest in a power plant you have electricity to run lights, power tools, etc. St. John encourages participants to think about this analogy when considering their projects as a means of building infrastructure for improving education.

Investments in infrastructures also bridge critical disjunctures. Airports connect regions, bridges connect two sides of a river, and you gain power when you connect disconnected places. Good infrastructure also has equitable access, and if you create an infrastructure only some people have access to, that is not a good infrastructure. It has to be reliable, offering high-quality, trusted resources. It allows for cumulative growth so that growth begets further growth. “We can think about NSF not as investing in short term projects,” St. John states, “but actually building the nation’s long-term infrastructure for improving STEM education.”

St. John observes that it was Doug Engelbart, a Professor of Emeritus at Stanford and the inventor of the keyboard mouse, who invented the improvement infrastructure idea and coined the phrase. Engelbart talks about “capability infrastructure,” which in schools is the instructional infrastructure that allows you to do instruction. Engelbart also says that organizations need an improvement infrastructure with a set of capacities and resources that helps them get better.

St. John cites a phrase Engelbart uses: “We need to get better at getting better.” The

purpose of NSF funding is not to get our education systems better, St. John notes, it is to get better at getting better. MSP projects and other investments are really aimed at getting better at getting better. He encourages participants to use this as a focal lens when thinking about their own projects: “How is my project helping us get better at getting better?” This is why NSF emphasizes the knowledge coming out of these MSPs, he points out. If you publish a paper about how to do professional development, you are helping somebody else get better at doing professional development.

Key Elements of an Improvement Infrastructure

- People - with expertise and mandate for improvement, linked in a community
- Ideas - about structures and processes of improvement
- Tools - resources, materials, processes for undertaking improvement.

St. John observes that before “One NSF,” the NSF logo used to say “We invest in people, ideas, and tools.” Not just random people, ideas, and tools, he points out, but people, ideas, and tools that can help improve the nation’s STEM education. That is human capital, knowledge capital, and design capital, across a knowledge capital network, which offers another way of thinking about this. “Unless we can document growth of people, ideas, and tools,” St. John says, “we are missing a big part of what the investment is doing.”

The Improvement Infrastructure

The foundational structure that gives a system the capacity to design, implement and sustain a process of ongoing improvements in the functioning of the system.

Just as infrastructure supports the operation, the improvement infrastructure supports the ongoing process of improvement, St. John notes. He refers to the following quote.

The Improvement Infrastructure

Engelbart argues that we... are still focused around projects and task forces with short-term expectations and short-term lifecycles, and have been too much in love with chasing after the latest tools and technologies... The most important activity we can do is to develop the improvement infrastructure... and to encourage and fund cross-functional "improvement communities" whose members work on common challenges to explicitly improve improvement... In essence, the human network, supported with a stable, sound technology network, is the way to get better at getting better.

- From *The 21st Century Intranet*, Jennifer Stone Gonzalez

There is inherently a problem with NSF's funding if NSF continues to fund short-term projects, St. John opines. The theory is that a short intervention will fix a problem, have an impact, and then the problem is handled and we can move on. With an infrastructure investment you don't invest for three years, you invest cumulatively over time. The infrastructure builds its capacity over time, you get

more and more working assets, and you can do more and more. "To what extent is NSF making investments that build this capacity, this linked set of capacities, for ongoing improvement?" St. John asks.

Another way of looking at this is Engelbart's levels. On Level A are the people doing the work (farmers tending crops, teachers teaching students). That is the functional level. On Level B are the people and resources concerned with helping Level A. In education, this is the education infrastructure including the administrators, curriculum developers, and textbook authors. On Level C are the people who help those on Level B help the people on Level A. That, says St. John, is where NSF's investments have to be. NSF does not have the scale to work on the system or even the improvement of the system. The scale of NSF work is down on Level C.

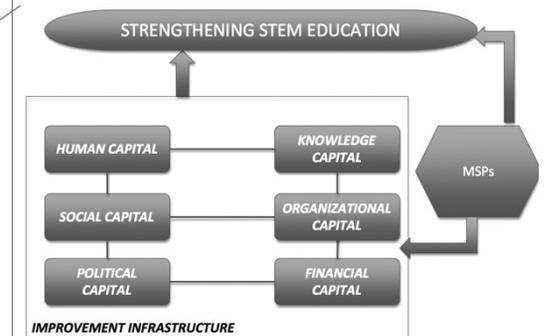
"When we take the educational capital we have been talking about and assemble it in a working way, we can begin to get an improvement infrastructure that is more permanent and is assembled in such a way that we can continue to do good work," St. John states. The Appalachian MSP serves as an example. In the last 15 years of investment in STEM education in Appalachia, there has been a growth in leadership, in organizations, and in social networks and connections. A number of things have come together to form an improvement infrastructure for Appalachia for STEM education. It is not entirely organized or funded, St. John reports. Nevertheless, this is a huge outcome of

Engelbart: Targeting the "Improvement Communities"

Engelbart's Levels A, B, C	Application to Farming	Application to Education
Level A -- people concerned with the process of doing work	Farmers growing crops	Teachers teaching students
Level B -- people concerned with the capabilities and conditions of Level A	Seed manufacturers, combine companies, irrigation people	Education infrastructure (Administrators, curriculum developers, textbook authors)
Level C -- people concerned with improving the work of Level B people and ultimately Level A work.	Seed researchers, combine developers, water engineers	Improvement infrastructure (Professional networks, R&D projects, developers of tools and resources)

Capital Assembled and Organized Becomes an Improvement Infrastructure.

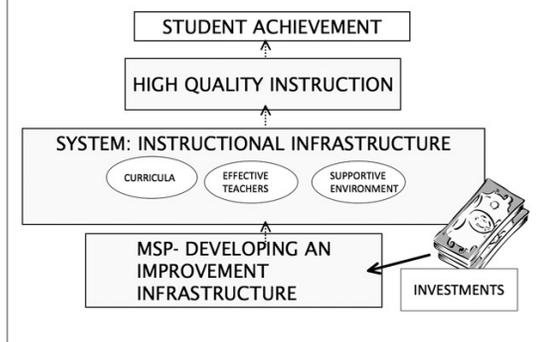
Assembling Capital To Create An "Improvement Infrastructure"



Assertion

The primary educational mission of NSF is to strengthen the nation's improvement infrastructure for STEM education.

NSF helps the nation get better at getting better.

The Nature of Investments Made in Educational Improvement

NSF funding which would be missed if you just looked at student achievement scores. “You would miss the fact that we are beginning to develop with multiple projects over time this infrastructure, which doesn’t do education but puts in place the ability to do ongoing improvement of education,” he states.

St. John has been arguing the assertion at left to NSF for years. Again, he reiterates, this is the lens evaluators should have: In what way are we building long-term capacity so that we can have ongoing improvement? It isn’t just what we are doing now, which is the expenditure; investment is in our long-term ability to keep improving.

There are probably five degrees of separation between student achievement and NSF’s investments, he notes. There is student achievement, which comes only in part from having access to high-quality instruction. There is not a one-to-one correspondence, St. John observes. Students spend 91% of their lives outside of school and many things go into a student’s ability to achieve. Access to high-quality instruction is just one element.

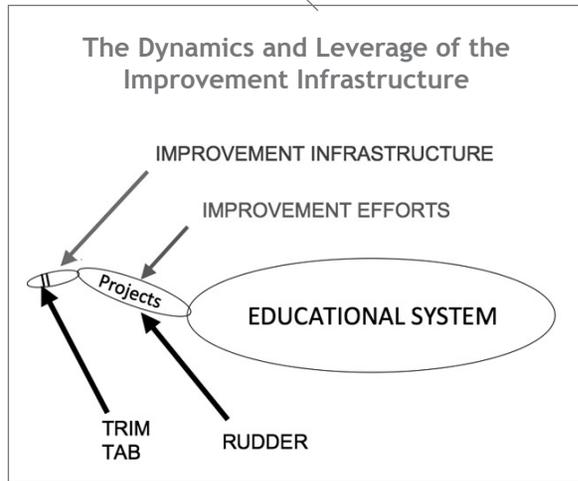
High-quality instruction, in a classroom or in a museum, comes in part, and only in part, from having a good system that supports it. This instructional infrastructure has to be strong, with good curriculum, well-prepared teachers, and a supportive policy environment. Together, these create a milieu, a supportive context for instruction, which leads to student achievement.

How do we get strong school systems? We need an educational improvement infrastructure, something that is almost entirely lacking in our country, St. John states. In other domains, such as aviation or pharmaceutical drugs, huge investments are made in the improvement infrastructure. In education we fund something from time to time to fix a problem, as opposed to having an ongoing infrastructure to continually help our instructional systems.

For example, St. John relates, the Maine PSP is trying over five years to develop an improvement infrastructure not just to do the work but to build capacity so that the districts and schools can continually support teachers and get better. If we do not evaluate that, St. John says, we are missing a huge part of what NSF can actually do. Evaluating student achievement is like saying, “My great, great granddaughters should be smarter because of what I’m doing now.” It is too far downstream and too difficult to make that assertion. We should be addressing a proximal goal, which is building capacity.

Below is yet another way of thinking about it, underscoring the importance of this point and of evaluators having this perspective. This is a trim tab analogy, St. John explains. On an airplane, the rudder turns the plane. When a school district has projects or undertakes an improvement effort, they are trying to turn the system, and you want districts to have the ability to turn the system. Some rudders have what is called a trim tab, which is a rudder on

a rudder. The trim tab turns the rudder and the rudder turns the system.



What we are really investing in is the trim tab, St. John explains. “We are investing in that leverage. If we can turn the local projects, then they can turn the system. Again, understanding the investment in terms of the capacity to help improve improvement by helping to improve local systems is really important.”

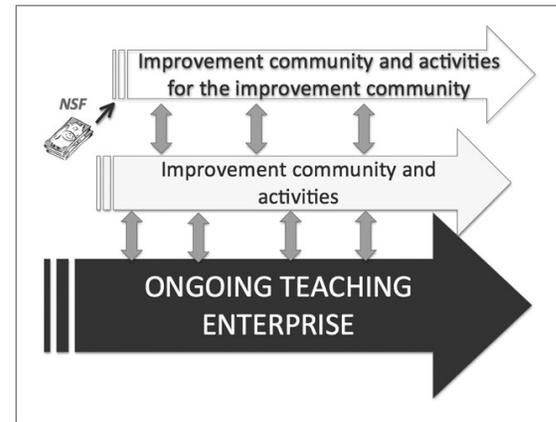
St. John offers one more graphic depiction of the argument (see sidebar). Teaching is an ongoing enterprise, he observes. We don’t say, “We had fourth grade last year, do we have to have a fourth grade again?” And yet we say that about projects: “We did a math project last year, but this year we are doing language arts.” Or, “We developed curriculum two years ago, do we have to do that again?” We have a notion of short-term, episodic fixes as opposed to ongoing capability to improve the system.

A better metaphor is gardening, St. John proposes. If we want a better garden, we don’t stop weeding it after one year, or water it once and then stop. Gardening is an ongoing enterprise. We need an ongoing enterprise locally that is constantly working with improvement. Currently, this is disappearing. We see district specialists and other local expertise disappearing. If we are serious about having local improvement infrastructures, we need to improve their improvement infrastructures. NSF has to build the national capacity to help local and regional efforts improve, and these have to be infrastructures so that they are continuous, not temporary.

How do we evaluate NSF investments using this lens? How do we assess whether NSF is building educational improvement capital, whether it building the improvement infrastructure? “We have started to do some of this work and to figure out how to do it,” St. John says, “but I welcome all of your attention to it because it is really embryonic, this measurement of capacity. We have spent billions of dollars measuring downstream achievement; we have spent very little money measuring capacity.”

There is a key question that evaluation should be asking: How are we building capital or an infrastructure for continued improvement? “That is the question that, as evaluators, I hope we will all be asking and trying to document,” St. John states. “Do we see MSPs doing that, and how?”

This involves entering a new domain. What are



Evaluating NSF investments using the lens of Educational Improvement Capital and The Improvement Infrastructure

Key Question
To what extent and in what ways does this project contribute to the capacity of the region to improve their STEM education and continue to improve it in the future?

The Process of Developing Evaluation Approaches to Measuring Educational Capital

- Definition - Conceptualize the likely forms of educational capital to be generated
- Instantiation - Look for multiple examples - extensive documentation of generation of capital
- Connoisseurship and expert judgment
- “Counting” - Some estimates of quantities or degree of development of capital
- Measurement - more careful documentation of capacities developed

the metrics for capacity? “The answer is, I don’t know,” St. John acknowledges, “but I can tell you how we might get there.” Conversations during this presentation are a beginning stab at trying to conceptualize and identify the kind of capital that MSPs are likely to be building. St. John says. He points to leadership development, with new leaders emerging in faculty positions, administrative positions and teacher leaders. Can we document that degree of growth of leadership, that human capital growth? There may be social capital, or networks of communities being created through collaboratives. Inverness Research is looking for those forms of capital that are likely to be generated and whether evaluators can begin to identify and document those.

A second step for evaluators in this process is instantiation. If we have a theory about something, like growing human capital and leadership development, we need to find examples and record those, St. John states. In Maine, Inverness Research does case studies and looks at teacher leaders and administrators, attempting to find examples of the kind of growth, participation, and involvement that offers convincing evidence that capital is being developed. “If we can’t find examples of something,” St. John observes, “it is probably not worth measuring. In fact, if we can’t find multiple examples to illustrate and instantiate it, it is probably not worth going after.”

The third step involves connoisseurship and expert judgment. When evaluating the Ap-

palachian Rural Systemic Initiative, Inverness Research applied a Capacity Framework to districts, looking at many different capacities important to math and science education and documenting their growth over time. This was done using an Expert Judgment Rating over time to fill out a Capacity Framework. St. John argues that sometimes the best instruments are the researchers: “The unbiased, systematic judgement of experts is a very powerful way to get at some of these rather complex and abstract ideas.”

The next step is counting. Is it possible to get some sense of things growing or developing in a way that we can begin to quantify it? “Finally we get to measurement, and the ability to measure capacity within a system or a district. I don’t know how to do that and I would be reluctant to say that I can measure it all,” St. John acknowledges, “but I can do the first four steps in various ways and at least begin to make the argument that the NSF investments are doing a lot for capacity building in the region.”

Inverness Research has a District Capacity Framework, which allows evaluators to ask

Example - A District Capacity Framework

- Leadership
- Instructional Improvement Capacities
- District Policies and Priorities
- Contextual Conditions That Influence The Development of a STEM Program
- Summary Judgments

questions of a district and look at capacities in a district for supporting, maintaining, and improving a science or math education program. “We are actually doing this work,” St. John reports. “Do I claim it’s perfect? No, but at least it is raising the issue. It is raising the visibility of the fact that these MSPs and other projects are creating capacity.”

In looking at how to evaluate infrastructure, it goes back to the criteria St. John talked about earlier. Are these criteria being met?

Criteria for Evaluating Investments in Infrastructure

- Empowering multiple functions
- Accessible and equitable usage
- Robustness, long term value
- Utilization
- Cost of usage
- Cumulative, growing
- Trust
- Multiple sources of funding

The best example of this, St. John notes, is the Inverness Research work with the National Writing Project (NWP), which they have been studying for 15 years. When that study began the NWP was, like everyone else, obsessed with student achievement and spent millions of dollars studying the question of whether the achievement of students in the classrooms of NWP teachers improved. They never talked in terms of building national capacity, St. John observes.

The NWP has 200 sites around the country involving university-district partnerships and school site partnerships. They are funded by the federal government and have been going for 30 years. They are now the largest, longest lasting professional development project in the history of mankind, St. John reports. They serve 100,000 teachers a year and they develop teacher leaders. There are 6,000 active teacher consultants acting as NWP consultants. Each of those 6,000 teacher consultants, on average, serves 1,400 teachers a year.

They have a system of capacity building through their social network, their social capital, through their leadership, and through their model, their organizational capital, St. John relates. They are now a national infrastructure for improving writing and have the ability to serve thousands of teachers, to continue to grow, to continue to improve their quality, and to work at very cost-effective levels. The federal expenditure per teacher in the NWP is on the order of six or seven dollars per teacher hour. For LSCs or MSPs it is on the order of \$40 per teacher hour. The NWP has become a very efficient cost sharing infrastructure, St. John notes.

Inverness Research has written this argument, attached data to it, and it is now an argument that the NWP uses with Congress, that this is a treasured national asset that Congress should continue to grow. “This is a feasibility case to me about the fact that we can take capacity and the idea of improvement infrastructure and use it as a focal point for evaluating invest-

Evaluating The National Writing Project

- Empowering multiple functions
- Accessible and equitable usage
- Robustness, long term value
- Utilization
- Cost of usage
- Cumulative, growing
- Trust
- Multiple sources of funding

ments and also for shaping future investments,” St. John states. “I think evaluators can be the trim tab, if you will, that helps foundations change their focus from being solely on student achievement to the array of benefits that come from these investments.”

Group Discussion

Funding Sustained Infrastructure Improvement

- It seems to me from what you’ve been describing that the U.S. Department of Education is a victim of its own rhetoric. The point of it is that you’ve got to find a solution to a problem, it’s not how to get better at getting better. It’s how to solve the problem of the fact that Johnny can’t read. How do you take these criteria and tell a meaningful story, with evidence, to a public and a Congress who are looking for a solution? In the NWP example, if I were a member of Congress I’d say, “If you’ve been around for 30 years and can’t sustain yourself by now, we ought to eliminate you.” In the capacity you described, I see a future that is building an answer to sustainability if we can tell it in a meaningful way, as opposed to people saying well, we know the way to sustain it is to look for another grant.

• Participant

- You’re saying that we need to make this argument in a way that is politically strong and addresses the criteria of cost-effectiveness,

sustainability, and so forth. One of the things about an infrastructure like the power grid, for example, is that it is funded on all levels—federal, state, and regional. Infrastructure is also paid for by consumers, so everybody chips in and the infrastructure is ongoing. In the case of the NWP, the state chips in, the districts chip in, so there are multiple supporters of this infrastructure.

This argument is one we continue to have to make. Things don’t get fixed. People don’t understand that you don’t have something without an infrastructure that makes it possible. When the Swine Flu hit, we couldn’t produce a vaccine quickly enough because we didn’t have the infrastructure in place. In Groton, where they have nuclear submarine building capability, they build one submarine a year, not because they need it but because they don’t want to lose the ability to build nuclear submarines, so they keep that infrastructure going.

We need to make a case that we need to build these infrastructures, and that our MSPs are building infrastructures that can be sustained and keep going. The whole thing requires a shift in our thinking about funding, from project “fix it” funding to ongoing infrastructure funding. You are right that it is going to take a different argument and different evidence to swing this around. • Mark St. John

- What you just said is, the evidence isn’t self-evident unless you say what you just said,

Participant discussions



unless you tell the rest of the story, and that is our job as evaluators. • Participant

- Correct, we need to have evidence aligned with the story. • Mark St. John

Inadvertently Creating an Infrastructure Deficit

- To bring this to a smaller scale, I'm in an MSP science educator project involving two elementary schools. The teachers love it and they get all of this free stuff, including technical assistance, but the funding is running out, so the person running this is going to go away at the end of the school year. In the meantime, the principal has shifted funds to other areas. If I were evaluating this project on the merits of the project alone I would say it is a great project but looking at the context, you see that now they have created a deficit in those two schools because that money has been shifted to other areas. Now I understand what you are saying about return on investment. You have to go outside and the specifics of just your project and look at what is going on all around it. • Participant
- That is very well said and it is exactly right. Most of the projects we look at do good work, but the question is larger. Does this create residual assets to continue the work? I don't care if the program is sustained, per se, but maybe there are people who go on to do things or new projects. If there is no legacy

at all, it is as if we were venture capitalists who started a firm, got it going, ran it for two years, and then said okay, that's it, let's do another start-up. This is about project churning. As evaluators we have to try to push more of a long-term effort but also recognize the capacities that continue after the project. The more we can do to put a lens on working assets, capacity building, etc., the more we can begin to say it is not enough to be a good project and do your little thing. That may be fine, but it's very short-term. • Mark St. John

A Portfolio of Arguments to Respond to the Social Environment

- I worry about the larger scale. What we are trying to accomplish isn't necessarily popular. We are in a very anti-intellectual environment in this country. Everything we do to demonstrate all the new knowledge we've created and all of the smarter, more independent-minded, creative individuals we've helped nurture runs up against a very strong milieu of: "Why should anybody go to college? The only reason to go to college is to get a certificate to get employment." There is a real, fundamental disconnect and NSF is caught in the middle, as are all of us. • Participant
- This is not happening independent of larger societal cross-currents. Part of this comes back to evaluators helping to make the case. Evaluators should be able to communicate a project to different audiences in different ways with different arguments and different

Participant discussions



Participant discussions

evidence, depending on the currency that audience uses. I don't disagree with you, I am saying that context then demands that evaluators and projects have multiple ways of explaining their value. It is not enough to simply have one line. Having a randomized experiment that shows student achievement might be helpful with some audiences and not others. We need a portfolio of arguments and evidence, and that's what evaluators can help to do, make the case, both at the initiative level and the project level. • Mark St. John

Language Driving Rhetoric

- In federal projects, when talking about money they always say you should supplement, not supplant. I'm thinking about the last few comments and the language we use. Instead of "supplement," we could use "capacity build," so it would be "capacity build instead of supplant." We could all start using a vocabulary that moves the rhetoric along in the general direction we feel it needs to go. I've always thought of "supplement" as being "capacity building," and in doing evaluations that's one of the things we are always looking at: Where are the capacity-building portions of this?"
- Regarding the logic model, you can let everybody stay in the boxes, but the arrows move around. The really hard bit is measuring where the arrows go. If we started doing that, kept the boxes that they're expecting but just moved the arrows around, maybe they wouldn't get so nervous. • Participant
- Your language point is really good. I was pleased in the LNC plenary session to hear Joan Ferrini-Mundy talking about human capital, and then Jim Spillane talking about infrastructure. For fifteen years I have been talking about capital and infrastructure. You finally begin to see it reflected not just because I did it, but because it comes with the territory. I think it's very important for evaluators to be leading agents, to bring a lens like this and say, "Hey, I am the steward of the investment and I think one of the things that comes from this investment is capacity building. Let's make sure we don't miss the capacity we're building." • Mark St. John
- In our project we are now calling professional development "human resource development" because it's developing the human resource of the environment, it's not just a piece you deliver to a school in a box. • Participant