SCALE

October 2, 2006

System-wide Change for All Learners and Educators

Evaluator's Report

Submitted to the National Science Foundation

This page left blank for print format

System-wide Change for All Learners and Educators

Evaluator's Report

Submitted to the National Science Foundation
October 2, 2006

The preparation of this report was supported by a grant from the National Science Foundation to the University of Wisconsin–Madison (EHR 0227016) for a Mathematics & Science Partnership project called the System-wide Change for All Learners and Educators (SCALE) Partnership. At UW-Madison, the SCALE project is housed in the Wisconsin Center for Education Research. The other five partners are the University of Pittsburgh, where the SCALE project is housed within the Learning Resources Development Center's Institute for Learning; the Los Angeles Unified School District; the Denver Public School District; the Providence Public School District; and the Madison Metropolitan School District. Any opinions, findings, or conclusions are those of the authors and do not necessarily reflect the views of the supporting agency

SCALE Evaluator's Report Table of Contents

Evaluation Summary - Andy Porter 1

District Case Studies – William H. Clune 12

Targeted Studies 16

1. SCALE Science Immersion Unit Summary: What we know about implementation - Bruce King 17

2. Studies on Science Immersion at the University of Pittsburgh - Chris Schunn 23

3. Madison East High School Curriculum Evaluation Study - Norman Webb 27

SCALE Quality Indicator System - Norman Webb 29

Case Studies of Institutions of Higher Education - Susan Millar 32

Other Work 35

Partnership Response to Evaluator's Summary 37

Publications and Reports: Citations and Abstracts 42

This page left blank for print format

Evaluator's Summary - Andy Porter

Introduction

The SCALE NSF Math Science Partnership has made a major commitment to research and evaluation of the work of the project and what it has accomplished. The purposes of the research and evaluation work are to:

- Provide formative and summative evaluation of SCALE work;
- Produce local knowledge of use to the partners; and
- Produce generalizable knowledge of use to the field.

During Year 4 of SCALE, there were five lines of research and evaluation work: (1) Building a Partnership (BP), directed by Susan Millar; (2) District Case Studies, directed by Bill Clune; (3) Targeted Studies, directed by Bruce King with components conducted by Chris Schunn and Norman Webb; (4) SCALE Quality Indicators System (SQIS) building an indicator system, directed by Norman Webb; and (5) Institutions of Higher Education (IHE) Case Studies, directed by Susan Millar. Andrew Porter directs the research and evaluation effort, and the goal manager is Sarah Mason.

SCALE's research and evaluation activity is referred to as the Research and Evaluation Team (RET). While the five lines of work are conducted concurrently and each with its separate team leader, each month the RET team meets as a whole via videoconference to present and critique each other's work. Each monthly meeting focuses on one major RET initiative. Materials are circulated in advance, brief presentations are made and in-depth discussion of the strengths and weaknesses of the work follows over the course of a two-hour session. These monthly meetings have served to not only strengthen the work in each of RET's five lines of research and evaluation activity, but have also helped to build community and ensure that each corner of the RET world understands what the other corners of the RET world are about and what they are learning from their work. [1]

Evaluation Activities and Findings

SCALE is making a major investment in research and evaluation work. The work is informing the direction and nature of SCALE efforts as the partnership matures. The lines of work are complementary, fitting together into a coherent program that addresses the breadth and depth of SCALE work. The following report is organized by line of work, each section summarizes last year's research activity and evaluation findings. Reports, documents, tools, and publications produced by the RET in 2005 and 2006 are listed in the *Publications and Reports: Citations and Abstracts* section of this report. Full copies of these materials may be found on SCALEnet at:
https://workspace.wcer.wisc.edu/gm/Year4PublicationsDashboard

[1] For ease of reference, the partner school districts will be referred to throughout this document as follows: LAUSD (Los Angeles Unified School District), DPS (Denver Public Schools), MMSD (Madison Metropolitan School District), and PPSD (Providence Public School District.

Highlights and findings of note include:

Building a Partnership –

- Preliminary data indicate that SCALE working groups addressing problems of "exploitation" are those that explicitly work on the challenges of broad implementation. SCALE working groups that focus on "exploration" challenges depend for their effectiveness on a loosely coupled organizational structure and on the presence of non-authoritarian "boundary-crosser" leadership that mobilizes the engagement of people with different viewpoints and expertise.

- A preliminary analysis of case studies designed to study K12-IHE partnership operations in four SCALE working groups reveal that co-constructing curriculum and professional development programming forged cross-institutional collaboration among members of one working group. Leaders developed social capital for the working group and its products by intentionally choosing or designing group processes and language that promoted knowledge and resource sharing. Leaders were also strategic in selecting knowledgeable, collaborative, and influential teachers, professors, and administrators to design and implement science curriculum and professional development. The leaders were also sensitive to the limitations of members and groups, and worked to ensure the group leveraged resources appropriately.

District Case Studies –

- The impact of the SCALE partnership on district policies of instructional guidance and distributed leadership at the top of the system has been significant. The districts that joined SCALE had theories of action (or strategic plans) that were a good match for the partnership theory. Indeed, the partnership theory was co-constructed with the districts and in many ways was a distillation of common elements. Excellent access to district leadership provided by the partnership fostered a working relationship and yielded rapid development of key elements of policy and organization. Two tasks are incomplete: successfully pushing guidance and support out to schools and classrooms, and building networks of sustainable leadership and support.

- At the SCALE Middle School Math Forum, collaboration of districts and experts converged on a central goal—improving the knowledge and skills of teachers around the curriculum as actually taught—and a set of designs (tools) for achieving this goal. Designs included formal professional development organized around big underlying ideas in mathematics, short curriculum units for students on the fundamental concepts present in every textbook, site-based professional development in which coaching becomes part of professional learning communities, inexpensive fast-turnaround formative assessment, and coaching organized around student work.

- A case study of interim assessments (IA) in PPSD concludes the IAs in mathematics have important strengths, but unresolved problems threaten long-term sustainability of the program. On the plus side, there was broad recognition among K-8 teachers of value or potential value of IAs for improving instruction

and support for the IAs remains strong at the central office level despite high turnover of district staff. But problems serious enough to threaten sustainability do exist. Major work is needed in the high schools on standards, assessments, and leadership. Equally important, the problem of how to find time for remedial instruction must be addressed because confusion about how to modify instruction undermines the central purpose of formative assessments.

- The approach being pursued in middle school science in LAUSD emphasized co-construction of curricula and professional development models as well as co-delivery of immersion professional development. The professional development model emphasized depth of teacher preparation (e.g., week-long institutes with substantial science content knowledge). In contrast, the theory of change in elementary science in LAUSD placed greater emphasis on breadth of reform—disseminating immersion units rapidly to relatively large numbers of teachers. This strategy reflects a decision in elementary science to emphasize breadth over depth in efforts to disseminate reform.

Targeted Studies –

- A MMSD immersion study shows that even in a high performing district such as MMSD, there was evidence of gains in content knowledge from this professional development, which was focused heavily on the implementation of the unit. The effect size was larger for design concepts than for science concepts, reflecting the higher initial state of knowledge about the science concepts.

- A study conducted by LAUSD Program Evaluation and Research Bureau (PERB) states that at the general level, the SCALE partnership is recognized for (1) development of immersion units, (2) professional development planning and delivery, and (3) collaboration with PERB on developing research tools. Out of five types of teacher training experiences that were observed and rated by PERB (meetings with instructional leadership team, grant-related professional development like California math-science partnership, immersion unit training, Math-Science-Tech Center workshops, and professional development delivered by local district instructional leaders), the effectiveness of the facilitator in guiding learning activities was rated the highest in immersion unit training (six events). Interviews with teachers receiving immersion unit training, showed that immersion implementation faced a number of hurdles including: limitations on teachers' time, lack of materials and resources, classroom management issues, testing demands, and scheduling constraints. Although not having sufficient time for an extended unit of study was the most cited complaint: "Many teachers were nonetheless very positive about the potential of the unit for their students."

- The findings for the University of Pittsburgh immersion units have continued to be largely positive. Researchers attribute the success in implementation to distributed teacher training rather than front-loaded training. In particular, professional development for the Pittsburgh unit consists of an initial exposure event for teacher recruitment, one workshop prior to teachers teaching the unit, and four workshops distributed over approximately an eight week period of time during which teachers are teaching the immersion unit. The results from the

Electronic Alarm unit study suggest that the systems design approach had superior performance in student knowledge gains on core science concepts, knowledge retention, and engagement when compared to the inquiry approach. The systems design approach was most helpful for low-achieving black students.

- The East High School, Madison, Wisconsin, mathematics curricula study was designed to provide information to teachers and district staff to make a decision among three types of curricula—a traditional curriculum, a moderate reform curriculum, and an integrated reform curriculum. After the second year of a three-year study, students in the moderate reform curriculum have had higher gain scores.

SCALE Quality Indicator System

- The focus of the SCALE Quality Indicator System (SQIS) for 2005-2006 has been to acquire achievement data for the current year from the four school districts, design and develop a data warehouse system that can be used to hold and process district data, and to produce reports on achievement data over a period of time from prior to SCALE and with SCALE.

- The mean achieve scores, as measured by a state assessment, by three of the four SCALE districts—MMSD, DPS, and LAUSD—have varied little from prior to 2003 (before SCALE) and after 2003 (with SCALE).
    - DPS grades 5, 6, and 10 mean mathematics scores had a small increase for 2004 and 2005. DPS grade 11 science scores for 2003, 2004, and 2005 are below the mean scores for 2002 and before.
    - MMSD mean mathematics scores for grades 4 and 8 had a drop in 2004, but then increased to above 2003 levels in 2005 after a general decline in mean scores from 2000 to 2003. MMSD grade 10 scores in mathematics have steadily declined since 2000. In science, the MMSD grade 4 mean scores have had a steady decline from 2000 to 2005, have been relatively flat for grade 10, and have declined for grade 8 until 2005. The mean achievement score for grade 8 science increased in 2005 after a steady declined from 2000. It should be noted that a large amount of SCALE efforts in MMSD has been directed towards middle grades science.
    - LAUSD mathematics mean scale scores for grades 4, 8, and 10 in general have declined for 2003 and 2004 compared to 2002 and before. Trend scores for LAUSD are difficult to interpret because a different test was administered in 2003 (CAT6). The science data is very lean.

IHE Case Studies

- *California State University Dominquez Hills (CSUDH) Case Study* - The study finds that one of the most important impacts that the Quality Educator Development (QED) and SCALE projects had through fall 2005 at CSUDH was the increased level of trust and collaboration between faculty in the College of Education and the science and math departments. This represents a significant shift from past history when faculty in the College of Education and the math and science departments were largely estranged from each other. Interviewees identified this improving relationship as a product of a number of the SCALE and

QED initiatives, including the professional development sessions for STEM faculty run by a faculty member from education, and joint participation in the development of science immersion units for LAUSD.

- *University of Wisconsin- Madison (UW-Madison) Case Study* - Through the *Math Masters* and *Immersion Unit* professional development programs for K-12 math and science teachers, SCALE is engaging STEM faculty in learning and modeling inquiry-based pedagogy, which is influencing the faculty's conception of their own teaching and of K-12 issues. Through the co-construction of professional development materials and the co-facilitation of the actual sessions, SCALE is introducing a new, more collaborative and mutually beneficial partnership between UW-Madison and MMSD. Finally, SCALE is also leading inter-departmental efforts to revise the pre-service math and science curriculum for elementary and middle school teacher candidates. However, the core SCALE strategies face significant barriers at UW-Madison due to an organizational climate that favors research over teaching and service, structural constraints for individual faculty, and pervasive tensions between STEM and education faculty.

Challenges and New Plans

As the work progresses, the RET continues to evaluate the scope and focus of work as a coherent program. There are tensions among the purposes the work is to serve. There is a challenge in meeting NSF reporting requirements and meeting the needs of all SCALE partners in districts and universities as they seek to realize the SCALE goals. There is tension between producing knowledge that is timely and formative, so that the work of SCALE implementers profit through reflection on what has already been accomplished, and work that produces generalizable knowledge of value to the field.

As for summative evaluation, one of the biggest challenges is to build causal arguments, the so-called attribution question. We wish to be able to attribute university and district actions and their subsequent effects to the SCALE interventions, yet the SCALE goals for university and district partners are not new to those institutions. How can SCALE research and evaluation distinguish between what SCALE has caused to happen from what would have happened whether or not SCALE existed?

As the SCALE interventions have matured and begun to be taken to scale, documenting the effects of these interventions becomes much more costly. The modest budget of targeted studies is not sufficient to keep up with the needs for summative evaluation. As a result, the targeted studies have been phased out and will not be continued in Year 5. The separately funded Gamoran study described in the report under "Other Work," is one example of the replacement for targeted studies. The level of work reflected in the Gamoran study of elementary school science immersion can only be pursued through separate substantial external support.

One final challenge is coordinating data collection. With five concurrent and closely articulated lines of work, data of interest to one effort is sometimes of interest to another. We have invented and put in place coordinating mechanisms that keep SCALE researchers from tripping over each other in their efforts to collect the data needed for

their work. The result has been jointly designed data collection efforts and shared data systems. The coordinating efforts have reduced the potential confusion and excess burden that would otherwise have resulted from redundant and overlapping data collection efforts. Especially in LAUSD, the coordination of data collection and securing of district permission received substantial attention. The result has been satisfactory to all parties involved.

Upcoming research and evaluation plans and activities include:

- At the end of five years, the Building a Partnership team will have produced a book that clarifies the nature and promise of the SCALE partnership with lessons learned that should guide the efforts of others who seek to reform education through building new and creative partnerships involving Research I universities, districts, and local IHEs.
- The district case studies team will have produced a book that documents the depth and breadth of SCALE-created change in district instructional guidance systems and the resulting effects on school and classroom practice.
- The indicator work will have documented how levels of student achievement are or are not changing from base line years prior to SCALE through SCALE's five years of activity in the four partner districts: LAUSD, DPS, PPSD and MMSD. The indicator system will not only document changes in student achievement and instructional practices, but will build new capacity in partner districts for analyzing value-added student achievement. Currently, only MMSD is engaging in value-added analytical work with SCALE.
- Implementation rates for SCALE immersion units are an important and pressing topic. More work needs to be done to establish how implementation can be uniformly high. The RET is tentatively considering a mini-conference on immersion to compare and contrast different approaches, to review the original concept paper for immersion design and delivery, and to attempt to learn what seems to be working and why. The conference would focus on the who, what, when, and where of different approaches to immersion as well as what has been learned about the characteristics of effective professional development over the course of the past one or two decades.
- The IHE case studies will have documented SCALE influences in local IHEs. Although change in institutions of higher education is notoriously slow and frequently not substantial enough to make a real difference; early evidence from SCALE work in the LAUSD area suggests that this disappointing history may have a brighter future.

Building a Partnership – Susan Millar

Year 4 progress is discussed under four headings:

1. SCALEviews
2. Mapping
3. Literature review on K12-IHE partnerships
4. Working Group Case Studies

.

All of the work described in these sections is designed to produce independent products and to provide data for use in the Building a Partnership (BP) book manuscript.

1. SCALEviews

The SCALEviews line of work (led by S. Millar) is pursuing the broad research question: *Why do K12-IHE partnerships form, and how do K12 – IHE partnership function?* Using SCALE as our source of information, we are reframing this question as follows: *How do SCALE participants express their expectations for, and understandings of, SCALE goals, strategies, processes, and outcomes?* Data on this question was gathered during interviews with 72 SCALE participants, starting in June 2005 and ending in April 2006. The interview period was much longer than anticipated due to difficulty gaining access to two SCALE districts. Interviewees were chosen on the basis that they are "key" people who help design and implement the work needed to accomplish SCALE goals. (Thirty of these 72 people also were interviewed in 2004-05.)

The 2005-06 interview questions were informed by a preliminary finding from the 2004 interviews, interpreted in light of ideas from complexity theory applied to organizations. This finding, which was presented at the NSF SCALE Site Visit in July 2005, is that there appears to be a relationship between the way SCALE working groups function and the degree to which the problems on which a group focuses are "exploratory" or "exploitative" in nature.

Some groups focus on "exploitation" problems. These are problems that experts know how to identify, define, and solve, where the knowledge necessary has been transformed into operating principles, where legitimized organizational procedures (templates) guide what to do, and where role authorizations guiding who should do what have been established. In addition, these are problems where the likely outcomes are known, based on prior implementation, where implementers can be assured that they can leverage resources effectively, and where success depends on broad implementation with fidelity. Preliminary data indicate that SCALE working groups addressing problems of "exploitation" are those that explicitly work on the challenges of broad implementation. These groups appear to be aware that authoritative action and relatively tightly coupled organizational structures are needed in order to protect system boundaries from disruption by outsiders, and to increase concentration of desirable resources and organizational efficiency.

Other groups focus on challenges that require "exploration." These are challenges that even experts cannot clearly define, let alone solve, and where operating principles are

only vaguely understood. Exploration approaches are appropriately used when the problems are long-term, widespread, or involve a looming disaster, where fast, reliable feedback can be used as a tool, where there is a low risk of catastrophe from the exploration. In these scenarios the risk is high, however the payoff also is high if a creative breakthrough is achieved. Preliminary data indicate that SCALE working groups that focus on exploration challenges depend for their effectiveness on a loosely coupled organizational structure and on the presence of non-authoritarian "boundary-crosser" leadership that mobilizes the engagement of people with different viewpoints and expertise. Other attributes of groups that work effectively in exploration mode are that they depend on "creative abrasion" to produce new knowledge and tools, that they know how to manage conflict while engaged in their work by establishing trust, and through use of respect for good argumentation based on evidence.

The 2005-06 interviews explored the issue of exploitation vs exploration in working group operation, thereby gathering many narratives about the relationships between the tasks on which working groups focus and the dynamics with which working groups function. Of note, the data on working group dynamics gathered in these interviews is much less intensive that that gathered in the Working Group (WG) Case Studies (see below).

During spring 2006, we began analysis of transcripts of these 72 interviews, plus 45 transcripts from the most useful of the 70 interviews conducted in 2004-05. A "big-bin" coding structure (major themes = input, process, and output) was developed during meetings that included all team members and sometimes Eric Osthoff. Using this coding structure, all 117 transcripts were then coded in analysis software (Nvivo) during summer 2006.

During summer 2006, we realized that to answer our research questions, we needed to include in our analysis process the stories that interviewees told about what they, their SCALE working groups, and their home organizations are getting (outcomes) from SCALE, and about the processes by which they are getting these outcomes. Hence, we developed a process for extracting these stories from our interview material. Aware that we do not have the resources to extract all the stories about the approximately 80 WGs described by our approximately 100 interviewees, we chose to select a sample of those WGs and individuals most responsible for pursuing SCALE goals, and most able to describe the typical (for SCALE) challenges encountered, and strategies used to solve these problems, and the most salient types of outcomes of these efforts.

By September 31, 2006, we will have drawn our sample of key WGs and individuals, using as our source of information the interview data coded in Nvivo and entered into SCALEbase (a relational database designed for use by both the RET and the SCALE Administrative Office). We intend that the WG sample will include at least 3 sponsored by each of the eight SCALE partners, and will not exceed a total of 40 WGs. We intend to select about 10 key individuals. As of the end of December 2006, we will have drafted summaries of each selected WG and key individuals, drawing on the interviewee's stories. The summaries will include main goals sought, challenges encountered, strategies used, and outcomes obtained. While drafting these summaries, we will have identified

gaps in the stories that interviewees provided, and have planned protocols (for interviews in early 2007) designed to help fill those gaps.

2. Working Group Mapping

The Working Group Mapping line of work (led by S. Millar) is pursuing this research question: *How is SCALE structured, and how does the structure change over time, to meet demands associated with K12-IHE systemic improvement of math and science teaching and learning?* New data addressing these questions was gathered during the 72 SCALEviews interviews conducted during June 2005 – April 2006, and entered into SCALEbase. The data entry and data cleaning process was very time-consuming because the database needed substantial modification as the data-entry process proceeded, the data are complex, and email and phone inquiries had to be made to obtain additional information on both SCALE participants and working groups. The emerging set of current SCALE working groups is being used by the SCALE PI as a resource for understanding how efforts to accomplish SCALE goals are being undertaken on the ground. During September – December 2006, Millar and Tran will have analyzed the SCALE working group data in SCALEbase, and will have represented these data in the form of organizational maps. These maps, along with descriptive analysis, will be presented in an informal report for SCALE leaders by the end of March 2007.

Outcomes of this "mapping" work are both methodological and substantive. The BP team's first organizational mapping report (September 2004) made evident that, while organizational charts are valuable because of their simplicity, and ability to represent change over time, they may be too simple to provide a useful representation of the lived organizations they attempt to model. "The first BP mapping report provided organizational maps, which depicted and analyzed SCALE's numerous partnership working groups, partially overcame this limitation. It also informed subsequent data gathering. Emerging data from the current mapping database (to be provided in a comprehensive form in spring 2007) show that the working groups within SCALE have changed substantially. For example, approximately 15 groups have multiple sponsoring organizations, compared to only two in spring 2004. In addition, many groups have been identified, as successors to earlier groups, as group tasks and participants successfully evolve, and many others have disappeared either because they completed their tasks or because, for various reasons, they did not function effectively. The key role that "boundary crossers" play in groups comprised of people from several partner organizations is notable. Among many other patterns, the study also is exploring the relationship between types of problems that groups focus on and the roles of boundary crossers within groups.

3. K12-IHE Partnership Literature Review

The K12-IHE Partnership Literature Review line of work (led by M. Clifford) encompasses theoretical and research articles on cross-organizational partnerships in education, with additional references to the healthcare and business fields. For this study, we are constructing an EndNote database. Currently, the database is populated by 73 abstracts. Findings from the draft literature review already have proven of great value in the development of the book plans, and have been used in all BP papers written and

presentations made during this period. By December 2006, Clifford will have made the EndNote database available to all WCER researchers, and, by request, to others. The literature review will be made available to researchers by April 2007.

Preliminary findings suggest that research literature on K12-IHE or other educational partnerships is limited primarily to program descriptions, whereas research in business and healthcare fields is more advanced. Second, we found that the partnership literature contains multiple, nuanced definitions of "partnership" and multiple similar terms to describe this type of organization, including "strategic alliance," "coalition," and "advice network." Although researchers conclude that partnerships appear to be increasing in number, the research lacks a common definition and language for the construct. Third, the review examined what we know about motivations to form partnership, what partnerships accomplish, and how partnerships achieve their results. We found that educational partnerships aimed to improve their constituent organizations' capacity to provide current services by aiming to improve teaching quality; train and retain more teachers; improve student performance; provide teacher professional development; develop curriculum; conduct action research or more formal studies; leverage grant money; streamline or ease costs of student matriculation; and explore solutions to persistent problems. In terms of outcomes, a common finding from the business research is that most partnerships fail to reach stated goals. Educational partnership outcomes have commonly used participation or participant satisfaction as success indicators, and while a few educational studies reported student and teacher learning outcomes, study methodologies raised attribution questions. With two exceptions, the examined research studies are based on one or a small number of cases. Finally, while some factors, such as "trust among partners" or "agreement on goals," appear to influence business, healthcare, and educational partnership success, we know little about what people do with resources in partnerships to develop and maintain success factors and achieve results.

4. Working Group Cross Case Study

The Working Group Cross Case Study line of work (led by M. Clifford) is pursuing the research question: *How and why do multi-institutional working groups contribute to build leadership capacity for math and science teaching and learning improvement?* Building upon literature review findings, the purpose of the study is to understand what people do with resources in partnership to develop and maintain capacity for math and science teaching improvement. This study involves a cross-case analysis of four case studies of SCALE working groups (two each in 2006 and 2007). The purpose of the working groups studies is to provide an in-depth, detailed account of K12-IHE partnership operations in an effort to explain why these partnerships form, how cross-institutional collaboration is developed, and how K12-IHE partnerships are positioned to lead change efforts within partnering institutions. The first two studies, focusing on the LAUSD Secondary Science Immersion Group (SSIG), and on the MMSD High School Equity Group, will have been completed by December 20, 2006.

Although the first two cases have not been completed, a preliminary analysis of the LAUSD Secondary Science Immersion Group was presented at the 2006 annual meeting of AERA. The LAUSD SSIG was formed by California State University professors, LAUSD teachers and administrators, and UW-Madison staff to design curriculum and

teacher professional development in support of “immersion” science. The SSIG was led by middle-level administrators, including the SCALE PIs and staff, who coordinated operations and mediated external pressures.

Two preliminary findings were explored in the American Education Research Association (AERA) paper. First, the case drew a distinction between the official and lived partnership organization. Groups of the people who became official participants within the SCALE SSIG worked together prior to SCALE on SCALE-related reform issues in their separate home institutions, and only knew of like-minded colleagues in other organizations. The SSIG built cross-institutional collaboration among many of these members by engaging them in co-constructing curriculum and professional development programming. Second, the case shows how leaders developed social capital for the working group and its products by intentionally choosing or designing group processes and language that promoted knowledge and resource sharing. In short, leaders built group processes that required cross-institutional collaboration and knowledge sharing among group members, and developed a common language within the group. SSIG leaders were also strategic in selecting knowledgeable, collaborative, and influential teachers, professors, and administrators to design and implement science curriculum and professional development. The leaders were also sensitive to the limitations of SSIG members and groups, and worked to ensure the group leveraged SSIG resources appropriately.

The preliminary analysis raised some interesting questions that subsequent cases and analyses will address. First is a question about “partnership” as an organization, and the role of cross-institutional collaboration in partnership. This case showed that many SSIG members were a part of the SCALE partnership in Year 1, but they did not collaborate cross-institutionally until the SSIG began to function. Micro-level analysis of participation suggests that partnerships, officially, may or may not involve co-construction or joint work. Second is the question about partnership leader roles and practices. The case shows that SSIG results from considerable leadership work, and is not a self-organized group. In this case, the leaders, working behind the scenes, used considerable knowledge and power to select and leverage human and financial resources to build the group. Our analysis of leadership tasks will be developed further in the planned cross-case analysis.

District Case Studies – William H. Clune

Year 4 progress will be discussed under five headings:

1. The 2006 AERA case studies symposium
2. Development then abandonment of two in-depth case studies in DPS
3. Report on and evaluation of the Middle School Math Forum
4. The in-depth study of interim assessments in PPSD
5. The in-depth study (or studies) of science immersion in LAUSD

The first three items are complete. The second two items are under way and hence will be discussed last.

1. The 2006 AERA case studies symposium

Five new papers derived from the previous panoramic case studies were presented in April, 2006, at the annual meeting of AERA. Case studies of DPS, LAUSD, MMSD, and PPS were presented. Clune wrote a cross-site synthesis. Comments were given by Andy Porter and Lauren Resnick. Abstracts of all papers are included in this report (see pgs. 40-42). This abstract of Clune's cross-site synthesis summarizes the findings:

> The impact of the SCALE partnership on district policies of instructional guidance and distributed leadership at the top of the system has been significant. The districts that joined SCALE had theories of action (or strategic plans) that were a good match for the partnership theory. Indeed, the partnership theory was co-constructed with the districts and in many ways was a distillation of common elements. Excellent access to district leadership provided by the partnership fostered a working relationship and yielded rapid development of key elements of policy and organization. Two tasks are incomplete: successfully pushing guidance and support out to schools and classrooms, and building networks of sustainable leadership and support.

2. Development then abandonment of two in-depth case studies in DPS

The in-depth case studies (see Year 4 Implementation Plan) got off to a false start with two proposals for in-depth case studies in Denver. Both studies were directed at meeting a need identified in the panoramic case studies for systematic data on how the district instructional guidance system influences instructional quality, a need articulated in this way by a DPS staff member: "We launch all these new curricula, we have all these programs of professional development, and we don't know anything about whether they are working." The first proposal was for development of an indicator system for evaluation and continuous improvement of district instructional guidance. This proposal was dropped as too ambitious for existing resources. The second case study proposal for a study of the implementation of two high school courses--Cognitive Tutor Algebra and Biology: A Human Approach--was abandoned in favor of the in-depth study of interim assessments in Providence (see below).

In-depth studies based on the previous theory of action in DPS (described in the panoramic case study) became impractical. When the new superintendent took office, Sally Mentor-Hay stepped down as Chief Academic Officer, and SCALE ceased to have a role in the core instructional guidance system. The previous theory of action built by Mentor-Hay (that the proposal was designed to test) was scrapped in favor of a new strategic plan. Both proposals were reviewed internally, reviewed by district staff (many of whom subsequently transferred out of the central office), and presented at RET videoconferences.

3. Report on and evaluation of the Middle School Math Forum

On December 11-12, 2005, SCALE sponsored a day-long forum on new ideas for middle school mathematics and algebra. Representatives of all four SCALE partner school districts participated along with expert consultants from the university partners UW-Madison, University of Pittsburgh, Institute for Learning (IFL), and CSUDH. A special presentation was made by Professor Uri Treisman, of the Dana Center at the University of Texas at Austin, who participated in the sessions.

The abstract of Clune's report summarizes the forum as follows:

> The collaboration of districts and experts converged on a central goal—improving the knowledge and skills of teachers around the curriculum as actually taught—and a set of designs (tools) for achieving this goal. Designs included formal professional development organized around big underlying ideas in mathematics, short curriculum units for students on the fundamental concepts present in every textbook, site-based professional development in which coaching becomes part of professional learning communities, inexpensive fast-turnaround formative assessment, and coaching organized around student work.

4. The in-depth study of interim assessments in PPSD

An in-depth case study ultimately adopted and successfully launched was of the implementation of interim assessments in Providence Public School District. Interim assessments are an important part of the SCALE theory of action under the monitoring dimension of Goal 1. In fact, the LAUSD version, called quarterly assessments (PPSD also uses a quarterly system), had been previously studied in LAUSD in what turned out to be the influential "focus group" report by Osthoff and Cantrell and subsequently evaluated by Andy Porter and Robert Linn. Interim assessments are extremely popular throughout the country at the present time, giving added importance to generalizable findings emerging from the PPSD study.

Clune and White are studying the implementation of interim assessments, their history, purpose, and technical characteristics, with a special focus on how the data are reported, used, and responded to, in light of their intended purpose of improving instruction. Document collection and interviews with district staff and teachers in six schools took place in Spring 2006. A wider sample of teachers will be interviewed in Fall, 2006. Preliminary findings from the first wave of data collection have been written and presented in cross-district video conference and were summarized as follows:

The overall conclusion is that the interim assessments in PPSD have important strengths, but unresolved problems threaten long-term sustainability of the program. On the plus side, there was broad recognition among K-8 teachers of value or potential value of IAs for improving instruction; and support for the IAs remains strong at the central office level despite high turnover of district staff. But problems serious enough to threaten sustainability do exist. Major work is needed in the high schools on standards, assessments, and leadership. Equally important, the problem of how to find time for remedial instruction must be addressed because confusion about how to modify instruction undermines the central purpose of formative assessments.

5. The in-depth study (or studies) of science immersion in LAUSD

The in-depth study of science immersion units in LAUSD developed in two stages under the leadership of Eric Osthoff: (1) a partnership-oriented study of the contrasting approaches used for design and implementation of elementary and middle school science; (2) a multi-faceted study of middle school science immersion presently in progress.

(1) The partnership-oriented study of the contrasting approaches used for design and implementation of elementary and middle school science. Initial research done to understand SCALE work in science in LAUSD was presented at the 2006 AERA annual convention. The paper focused on similarities and differences in how SCALE was affecting science instructional leadership, instructional guidance, and teaching and learning in elementary and secondary science. As of April 2006, we found that the approach being pursued in middle school science emphasized co-construction of curricula and professional development models as well as co-delivery of immersion professional development. The professional development model emphasized depth of teacher preparation (e.g., week-long institutes with substantial science content knowledge).

In contrast, the theory of change in elementary science placed greater emphasis on breadth of reform—disseminating immersion units rapidly to relatively large numbers of teachers. Although some LAUSD elementary teachers have participated in week long institutes through SCALE and QED, the district has on its own, without higher education co-facilitation, turned primarily to briefer two-day initial workshops for teachers in which science content knowledge supplementation is not systematically or extensively pursued. This strategy reflects a decision in elementary science to emphasize breadth over depth in efforts to disseminate reform. Over time it appears the elementary and secondary immersion implementation strategies may now be in the process of becoming more convergent. For example, elementary teachers are getting district-led follow-up workshops, though it appears new science content knowledge for teachers is still not heavily emphasized.

(2) A multi-faceted study of middle school science immersion. Going forward, the in-depth case study of science immersion in LAUSD will consist of a study of the work in middle school science. The study includes documentation of middle school immersion professional development design and delivery, teacher implementation of immersion units in classrooms, effects on student learning, and system support for immersion

teaching and learning at all levels (e.g., school, local district, central office, higher education). Because of the follow-through to classroom instruction and student achievement, the study is a hybrid of what previously had been considered separate categories in the RET design -- the in-depth case study and the targeted study-- the latter being defined by its focus on instruction and student achievement. In fact, the new study incorporates all of what had previously been done by Bruce King in his targeted study of science immersion in LAUSD.

The middle school immersion study methodology is multi-faceted. The focus of the study is on Grade 6 science immersion (Plate Tectonics). Three teacher institutes were conducted in the summer of 2006. RET researchers observed all three institutes to document delivery. The following instruments/research activities were also conducted with institute participants:

1. Administration of institute science content knowledge pre- and post-test (developed under NSF RETA grant by Horizon Research).
2. Administration of 2-part survey: (a) Teacher demographics, professional preparation, and measures of teachers' local school capacity for supporting inquiry science teaching and learning, and (b) A version of the *Survey of Enacted Curriculum* that has been customized to the content domain of the Plate Tectonics immersion unit. Institute participants take this survey at the institute to report on the instructional content (topics and cognitive demand) of the part of the curriculum devoted to plate tectonics as they taught it in SY 2005-06. They will take the survey again once they have completed the unit in SY 2006-07. This will allow assessing the impact of the immersion unit on classroom teaching.
3. Focus groups to assess teachers' institute experiences.

The following research activities will be conducted subsequent to the institutes:

1. 2-3 classroom observations of each of 30 randomly selected institute participants.
2. Interviews with all observed teachers.
3. Collection and analysis of student work samples for all observed classrooms.
4. Post-implementation survey of classroom content with all institute participants.
5. Analysis of district quarterly benchmark assessments to compare the achievement of students in classrooms of teachers who have taken the immersion institute to student in classrooms with teachers who have not.
6. Interviews with principals, local district and central office administrators to learn about district support for and obstacles to broad and deep implementation of middle school science immersion in LAUSD.

Part A of the teacher survey was designed in conjunction with Adam Gamoran, PI of an NSF Teacher Professional Continuum (TPC) study of elementary immersion in LAUSD. Using the same instrument will permit comparing school capacity issues in LAUSD across the elementary and secondary levels to better understand the range of organizational factors affecting effective implementation of immersion instruction. The classroom observation instrument for the middle school immersion study is the same instrument being used by LAUSD PERB researchers and the TPC study. This too facilitates comparing immersion initiative effects between elementary and secondary levels.

Targeted Studies

The targeted studies line of RET/SCALE work was designed to conduct timely formative assessment on emerging SCALE products was originally led by Norm Webb, with leadership taken over two years ago by Bruce King. In what follows, King provides a summary of studies done and lessons learned for SCALE science immersion as developed and delivered by the UW-Madison team, sometimes joined by the QED work at CSUDH.

Chris Schunn has also been developing and testing science immersion and his progress is reported in this section on targeted studies immediately following the reporting of Bruce King's work. The section on targeted studies ends with an update on Norm Webb's study comparing the effect on student achievement of alternative curriculum for high school mathematics.

Year 4 progress will be discussed under three headings:

1. SCALE Science Immersion Unit Summary: What we know about implementation
2. Studies on Science Immersion at the University of Pittsburgh
3. Madison East High School Curriculum Evaluation Study 2005-2006

1. SCALE Science Immersion Unit Summary: What we know about implementation - Bruce King

This report summarizes findings on implementation from eight targeted studies of SCALE immersion units. Tables 1-4 summarize the immersion units developed by SCALE and information on the related professional development. Headings 5- 8 provide summaries of studies of immersion implementation conducted in Pittsburgh Public Schools, and two SCALE districts, MMSD and LAUSD. The report is organized around the following headings:

(1) Table 1: Units developed for MMSD and/or LAUSD, Pittsburgh
(2) Table 2: MMSD Professional Development
(3) Table 3: LAUSD Professional Development
(4) Table 4: Pittsburgh Professional Development
(5) Electrical Alarm System Study, Pittsburgh Schools
(6) Electricity and Magnetism Field Study, MMSD
(7) District Science Plan Evaluation, LAUSD-PERB
(8) Student Work Samples from Variation and Natural Selection, two 7th grade classes, LAUSD, Fall 2005

(1) Table 1: Units developed for MMSD and/or LAUSD, Pittsburgh

Grade	Science Strand	**Immersion Unit**
Kindergarten	Life Science	*Analyzing Animals*
Grade 3	Life Science	*Structures of Life* *Investigating Responses*
Grade 4	Physical Science	*Electricity and Magnetism* *Rot It Right*
Grade 5	Earth Science	*Weather Forces and Prediction*
Grade 6	Life Science Earth Science	*Diversity of Life* *Plate Tectonics*
Grade 7	Life Science Earth Science	*Variation and Natural Selection* *Exploring Earth's Landforms*
Grade 8	Physical Science	*Density and Buoyancy* *Electrical Alarm System (University of Pittsburgh)*
Grade 9	Earth Science	*Global Warming*
Grades 9-12	Integrated	*Climate and Ecosystems*

(2) *Table 2: MMSD Professional Development*

Elementary Immersion Unit, FOSS, and Science Scope and Sequence Training

Grades K-5	Number of Attending Teachers	Total Hours*
14 workshops	209	2402

Secondary Science training (may or may not have involved Immersion Units)

Grades 6-12	Number of Attending Teachers	Total Hours*
Middle School, 5 workshops	113	1659
High School, 4 workshops	85	466

*Combined hours of all teachers; individual workshops varied (see *MMSD professional development summary by type of PD*, J. Watson).

(3) Table 3: LAUSD Professional Development

Immersion Unit Training

Grade 4-8	Number of Attending Teachers	Total Hours*
QED/SCALE Science Institutes	157	~4700
Other	? no data	? no data

*Approximately 30 hours per teacher who participated in one of 7 five-day institutes (source, 2005 Science Institute Study: QED Project).

(4) Table 4: Pittsburgh Professional Development

Immersion Unit Training

Grade 8	Number of Attending Teachers	Total Hours*
Electrical Alarm System, 5 sessions throughout implementation	13	260

*Combined hours of all teachers (see *Evaluating the impact of a facilitated learning community approach to professional development on student achievement*, Doppelt, Y. et al. paper for NARST Annual Meeting, April 2006).

(5) Electrical Alarm System Study, Pittsburgh Schools
Schunn and colleagues reported on a study that compared student learning on traditional scripted science inquiry and on design-based, systems approach of the Grade 8 unit, Electrical Alarm System: Design, Construction, and Reflection. Scripted inquiry was implemented in 20 classes involving 5 teachers and 466 students, and the Electrical Alarm System unit was implemented in 26 science classes involving 10 teachers and 587 students. The results suggest that the systems design approach had superior performance in student knowledge gains on core science concepts, knowledge retention, and engagement when compared to the inquiry approach. The systems design approach was most helpful for low-achieving black students. See Mehalik, M.; Doppelt, Y.; Silk, E.; Schunn, C. (University of Pittsburgh). ***Middle-School Science Through Design-Based Learning versus Scripted Inquiry: Better Overall Science Concept Learning and Equity Gap Reduction*** .

(6) Electricity and Magnetism Field Study, MMSD
King and colleagues reported on a field test of the Grade 4 Electricity and Magnetism unit involving three teachers at two schools in MMSD who implemented the unit in their classes in late Spring 2005. Teachers liked the training for the unit and agreed that the unit was well designed, but they agreed that they did not have sufficient time to complete the unit, and none did. The unit did not seem to immerse students (or teachers) in the science content as much or as well as it could. This finding was reflected both in observed lessons and overall scores on the two assessments of student learning (scored using the rubrics provided in the unit). In terms of equity of outcomes, the different scores by student groups were disappointing (white students did better than students of color, English language learners, and students on free or reduced lunch).

Teachers seemed to struggle with the intersection of student investigation (e.g., making the light bulb light) and providing content knowledge to help them make sense of their investigations. Student investigation activity dominated the observed lessons. The one exception was the substitute teacher who did much more didactic instruction. But overall, students did not grapple with complex scientific explanations of their investigations and this showed in their work on the assessments.

A few implications were noted. First with regard to content knowledge, equipping teachers with more complex understandings of the content seems essential. This may include having them critically consider exemplars of written explanations that show in-depth understanding and of classroom discourse that gets to a fairly deep level. Second, the unit itself and teachers in implementation need to structure inquiry activities that tap most features of inquiry. Putting the teachers in the role of students and having them experience successful inquiry lessons may help. Finally, how can these two be connected? That is, both initial training and ongoing support need to attend to linking content and process. In what ways can inquiry provide the knowledge to students so they can develop elaborated scientific explanations? How can teachers orchestrate lessons in ways that have students step in and out of inquiry to acquire content knowledge as needed? The field test of the Electricity and Magnetism unit suggests that this might be one of the critical issues for immersion development, teacher professional development, and implementation. See King, M.B., Davis, D. & Mast, G. Electricity and Magnetism Immersion Unit: Summary of Field Test. 2006

(7) District Science Plan Evaluation, LAUSD-PERB
The Kelly and Rickles[2] report has a number of insights about immersion implementation. Note that the report covers the 2004-05 school year, with the grade 4 and grade 7 immersion units "in the field." While PERB collected data on immersion implementation in grade 7 (*Variation and Natural Selection*) in 2005-06, there was no analysis reported here.

At a general level, the SCALE partnership is recognized for (1) development of immersion units, (2) professional development planning and delivery, and (3) collaboration with PERB on developing research tools. The authors note, "You will see how the immersion unit and associated professional development events consistently show up as the strongest models of inquiry-based instruction in science" (p. 54).

Out of five types of teacher training experiences that were observed and rated by PERB (meetings with instructional leadership team, grant-related PD like CA math-science partnership, immersion unit training, Math-Science-Tech Center workshops, and PD delivered by local district instructional leaders), the effectiveness of the facilitator in guiding learning activities was rated the highest in immersion unit training (6 events). Teachers in immersion unit training were invited from middle schools from the lower half of the School Characteristic Index, LAUSD's measure of overall school performance.

Interviews with "treated" teachers (those receiving immersion unit training) showed that immersion implementation faced a number of hurdles including: limitations on teachers' time, lack of materials and resources, classroom management issues, testing demands, and scheduling constraints. Not having sufficient time for an extended unit of study was the most cited complaint. Several teachers also noted that the unit only covered a portion of the standards they are supposed to teach. 31% of teachers receiving training on *Variation and Natural Selection* did not teach any part of the unit, 58% implemented it "minimally." More training and ongoing training were recommended. The authors state that, "many teachers were nonetheless very positive about the potential of the unit for their students" (p. 107).

Seventh grade teachers who received immersion unit training (treated) and teachers who did not received immersion unit training (untreated) were observed when teaching content focused on science standards addressed in the immersion unit. Ninety-six percent of lessons of treated teachers addressed one or more of the evolution and investigation/experimentation standards compared to 36% of lessons of untreated teachers. However, untreated teachers addressed genetics and plant physiology standards more than treated teachers (73% of lessons compared to 38%). Treated teachers used

[2] Kelly, K. and Rickles, J. (2006). *Baseline Report: The District Science Plan Evaluation, 2004-05.* Los Angeles, CA: Los Angeles Unified School District, Planning, Assessment and Research Branch. A draft copy of this report was shared with SCALE researchers for informational purposes. As of September 30, 2006, the report has not been publicly released by LAUSD.

more whole-class discussions and student-led activities than did untreated teachers. But among treated teachers implementing the immersion unit lessons, observations showed that they did not incorporate all features of inquiry (0% had students communicate and justify explanations or evaluate explanations based on understanding scientific content) and their lessons failed to reach higher levels of cognitive challenge (0% had students analyze information or make connections or apply concepts). About 60% of observed instructional time was rated high in student engagement for both treated and untreated teachers. Teachers implementing the unit who were interviewed expressed having difficulty completing the unit and that only parts of the last steps of the unit were completed.

Measures of student performance related to the *Variation and Natural Selection* immersion unit included depth of understanding demonstrated during observed immersion unit lessons and results on periodic science assessments. In lessons, students showed complex understanding of content in 21% of treated teachers' lessons and 0% of untreated teachers' lessons.

For the periodic science assessments, immersion implementation was severely limited, scheduling required testing before training and implementation in some schools, and results were reported by clusters of schools. Thus they do not seem to advance understanding of the impact of actual implementation of immersion on student learning. As reported, overall assessment results showed students in schools with untreated teachers outperformed students in schools with treated teachers on all tested items, as well as on just those items related to standards covered by the immersion unit. But it doesn't appear that results were disaggregated for students of treated and implementing *teachers*. These comparisons also do not account for the fact that schools with treated teachers were in the lower half of the School Characteristic Index. The regression discontinuity analysis compared predicted and actual periodic assessment results of students in 25 schools with treated teachers and 25 schools with untreated teachers (these 50 schools were closest to the School Characteristic Index median, with the 25 schools with treated teachers below the median and the 25 schools with untreated teachers above the median). The analysis showed students in schools with treated teachers had a 1% advantage over students in schools with untreated teachers on the total percent correct score, and a 2% advantage on test items related to standards addressed in the immersion unit. The authors state this analysis should be considered illustrative of the methodology and suggestive only, in part because some students in schools with treated teachers were tested *before* teachers received immersion training and any implementation.
From "Baseline Report: The District Science Plan Evaluation, 2004-05," March 9, 2006.

(8) Student Work Samples from Variation and Natural Selection, two 7th grade classes, LAUSD, Fall 2005
Although the implementation phase of the QED/SCALE Institute Study did not materialize in 2005-06, PERB did collect data on implementation from some grade 7 teachers who participated in the summer 2005 institute on *Variation and Natural Selection*. From this, PERB sent us two classroom sets of anonymous student work samples on one assessment from the immersion unit (Step 6). There were no other data on demographics or implementation, but we did proceed with scoring the work samples. The criteria used for rating the quality of student work were Scientific Analysis, In-depth

Understanding of Disciplinary Concepts, and Elaborated Scientific Communication. The first two criteria—scientific analysis and in-depth understanding of disciplinary concepts—are consistent with the kinds of cognitive work students should do when involved in many features of scientific inquiry, and should tap the quality of students' learning from inquiry activities and lessons. The third criterion—elaborated scientific communication—corresponds directly to the fifth feature of inquiry—learners communicate and justify explanations. Note that we did not receive feedback on the scoring criteria from Goal 2 staff. See King, M.B. Standards and Scoring Criteria for Student Work on Science Immersion. February 2006

Results on Step 6 Assessments

Standards (possible scores on each, 1-4)	Mean	SD
Scientific analysis	2.12	0.86
In-depth understanding of disciplinary concepts	1.90	0.63
Elaborated scientific communication	1.60	0.74
Total (range 3-12)	5.63	1.97

Scores tended to be low but there were no other contextual data, such as quality of teacher's implementation of immersion lessons, to make any interpretations. The scoring criteria need further development. Dalelia Davis did the initial scoring and reported that the first standard, scientific analysis was the hardest to score in ways that were consistent from student to student. The categories for scores of 2 and 1 were not easily distinguishable, except when the student showed only the answer and the answer was incorrect. The third standard, elaborated scientific communication was the easiest to score based on the descriptions provided as well as the nature of the standard. It took approximately two hours and fifteen minutes to score 53 student work samples the first time.

2. Studies on Science Immersion at the University of Pittsburgh - Chris Schunn

Chris Schunn has conducted a number of studies of the implementation and effects of electrical alarm system immersion unit. In last year's report, results of effects on student achievement were reported. These results were positive and appeared to also result in a reduction in the achievement gap between white and black students.

The findings for the University of Pittsburgh immersion units have continued to be largely positive. Of those teachers who say they will implement the unit, virtually all of them do. This contrasts with the early implementation results of other SCALE immersion units, as described in Bruce King's summary above. These findings led to speculations by the RET as to what characteristics distinguish the two immersion approaches and teacher training in support of the units that might explain the differences in implementation rate. The following is a list of hypotheses:

- The unit/text is district adopted.
- Other forms of central office encouragement to use the unit/text.
- Perceived degree of alignment with state/district content standards and testing (coherence).
- The teachers interest in covering the unit/text content.
- Front loaded versus distributed professional development (duration).
- Consistency with immersion unit characteristics.
- Amount of material for teachers to use.
- Amount of material for students to use.
- Other forms of classroom support (e.g. coaches).
- Number of hours of professional development (duration).
- Teacher pay for participation/making pay contingent.
- Content focus of the professional development.
- Collective participation during professional development.
- Active learning in professional development.

Chris Schunn believes that the key to his success in implementation of his unit is that the teacher training is distributed rather than front-loaded. In particular, professional development for the University of Pittsburgh unit consists of an initial exposure event for teacher recruitment, one workshop prior to teachers teaching the unit, and four workshops distributed over approximately an eight week period of time during which teachers are teaching the immersion unit. In contrast, the other SCALE immersion units' training consists of a five-day institute up-front (all before teachers actually teach the unit) with two half-day follow-up sessions during the semester the teacher's implement the unit.

1) A study of the impact of workshops on teacher learning---this was done in the context of work with 8th grade teachers in MMSD.

One of the goals of immersion unit professional development is to provide teachers with an opportunity to strengthen their science content knowledge, but in a way that provides a good model of pedagogy and does not threaten them as relative experts on the science

topics they teach. Schunn, Mehalik, Silk, and Doppelt tested this impact of immersion unit professional development on teacher science content knowledge in the context of their summer 3-day immersion workshop with 8th grade teachers of Madison Metropolitan School District. Thirteen teachers participated in the workshop, and completed pre and post-tests that were embedded into the immersion unit as they engaged with the immersion unit as learners. Additionally, teachers used pseudonyms of their own choosing on the tests to further reduce any perception of threat associated with being tested on content knowledge. Even in a high performing district like MMSD, there was evidence of gains in content knowledge from this professional development that was focused heavily on the implementation of the unit. The effect size was larger for design concepts than for science concepts, reflecting the higher initial state of knowledge about the science concepts.

(2) Ongoing analyses of the causal reasoning ability of students working with the high school chemistry immersion unit.

Grasping scientific phenomena includes more than the mere memorization of scientific facts. It requires the creation of an appropriate model of the important aspects of the phenomena and how these aspects are related. Many, but not all, of these relations are causal. We are attempting to understand how students acquire knowledge of these causal relations and how this knowledge is expressed. Furthermore, we are interested in how students' everyday notions of cause and effect and/or their everyday observations of scientific phenomena might constrain this process. Very recently, we collected extensive interviews and conceptual maps from 30 high school students after completing our design based learning curriculum in high school chemistry (10th, 11th, and 12th graders). This curriculum requires students to reason through the causal model of a scientific phenomenon. Other non-design based curricula that require this sort of reasoning seem to result in improved understanding of that scientific model and other general scientific principles related to that model (e.g., Reiser, 2004; Reiser, Tabak, Sandoval, Smith, Steinmuller, & Leone, 2001; Sandoval & Reiser, 1997; White, 1993; White & Frederiksen, 1998, 2005). However, little is still known about the nature and complexity of these causal connections or how this complexity might be related to student performance on independent concept assessments.

During individual interviews the students created conceptual maps that represent 'why some reactions get hot and some reactions get cold'. Initially, students selected cards, from a set of 40 that could be used to explain the previous question. Students were given the option to create their own card if they had an idea that did not exist. After selecting the relevant cards, students were asked to place those cards on a paper in a way that made sense to them. Afterwards, students drew connections between these cards to indicate how these cards were related to each other. Currently, we are analyzing the data from these conceptual maps. Our initial analyses indicate that students include more macroscopic terms than microscopic, with the number of microscopic terms increasing with student understanding. Further analyses will be conducted to investigate how the number and type of connections is related to understanding of the phenomena. Furthermore, we plan to investigate whether these connections are relational or causal. This is our first study in this area. Additional studies will explore how these conceptual maps evolve during learning.

(3) Investigating the Tacit Problem-Solving Strategies of Novice Designers

This study is a first step at exploring the problem solving strategies that students use, and the relationship of these designer problem-solving strategies to students' learning of science when engaged in design-based curricular activities. The purpose of this study is to investigate the problem-solving strategies of students as they attempted to design a solution to a novel design problem. The following research question is of primary interest in this study:

> What are the problem-solving strategies used by students as they engage in solving novel design problems?
>
> Did participation in the design-based curriculum unit influence the problem-solving strategies students used to solve the novel design problems?

Preliminary analysis of the Earthquake task video shows that students followed a wide range of unique solution paths in solving the design problem. Students constructed an average of 5 designs per session (M=5.3, SD=3.2). Of their design attempts students created an average of 3 successful structures (M= 3.5, SD=2.5). Regarding the strategies students used when designing, the add-on strategy was most prevalent with students using this strategy an average of 52% of the time. Data analysis of the Alarm System task is still underway.

(4) Evaluating a Design-based Learning Curriculum in Terms of Students' Science Reasoning Gains.

The objective of this research was to evaluate a reform science unit that the University of Pittsburgh immersion tam has been developing and researching over the past few years in terms of measures that will be sufficient for the accountability structures put in place by NCLB. The context for this research was the Electrical Alarm System unit, which is a design-based learning unit (Doppelt, Mehalik & Schunn, 2005) focusing on the teaching of core electricity concepts to eighth graders through the design of an alarm system that meets an everyday need (e.g., a locker alarm to inform me if someone breaks into my school locker). This reform curriculum was designed to supplement the first four-to-six weeks of instruction in an established scripted-inquiry curriculum by incorporating the open-ended design project as a launching pad for the semester-long study of electronics (Schunn et. al., 2004). The unit has been evaluated previously in terms of its impact on students' learning of the science content relative to students in the scripted-inquiry curriculum alone. The design-based unit was found to significantly improve students' learning of the science, especially for traditionally disadvantaged students (Mehalik, Doppelt & Schunn, 2005).

The findings support the claim that the Alarm System unit does an effective job at helping students to acquire science-reasoning skills when measured on validated standardized test items. These results occurred in the context of an authentic, engineering-design project that has only a secondary focus on scientific inquiry. In addition, the design-based unit appears to be effective even in classrooms comprised mostly of students from a low socio-economic background where an existing hands-on

elementary and middle school curriculum had not yet had much of an effect on their reasoning skills (cf., Pine, 2006).

The results are promising in that they provide evidence to justify the use of reform science curriculum in ways that are valued not only by communities of researchers, but also by education administrators and practitioners. District personnel, principals, and classroom teachers all have legitimate concerns that the reform science curricula that they adopt and implement ought to have observable impact on the types of assessments to which they are held accountable. Research-based curricula are often of high quality, as reflected in the alternative and contextualized assessments that the researchers/designers employ. Work such as the current project help establish whether reform curricula also impact students' performance on the standardized assessments, which is important because it may be that those standardized assessments are actually harder for students, especially for students of traditionally disadvantaged populations.

Curriculum developers need to justify the increased time and resources spent on participating in open-ended science activities when that almost always implies that less time will be spent covering the content in the traditional sense. Design-based units have the further obligation to look beyond the learning of engineering and technology skills to target science content and inquiry skills in order to be considered a practical alternative to inquiry-based units, both of which are consistent with science reform efforts. Articulating and documenting the advantage of design-based units for low-SES populations in particular will also serve to bolster design as a viable alternative for helping to create systemic change in science education.

3. Madison East High School Curriculum Evaluation Study - Norman Webb

The Madison East High School mathematics study is a cooperative study between SCALE and the Madison Metropolitan School District (MMSD). SCALE staff has worked cooperatively with MMSD district staff to conduct a study to produce data that can inform the district and high school staff on the effectiveness of different high school curricula and can be used to make decisions about what curricula should be used. This action research and data decision making intervention also is a SCALE targeted study of student achievement over three years.

Over three years, beginning in Fall 2003 with students enrolled in first year of high school mathematics (Algebra I and Integrated Mathematics I), students were administered achievement tests at the beginning of the school year and at the end of the school year. In the second year of the study, the 2004-2005 school year, two cohorts of students were given tests at the beginning and end of the school year—those enrolled in first year of high school mathematics and those enrolled in geometry or Integrated Mathematics II. In 2005-2006, the third year of the study, pre and posttests were administered to three cohorts of students—those in first year of high school mathematics, those in the second year of high school mathematics, and those in Algebra II, precalculus, statistics, or Integrated Mathematics III.

In 2005-2006, the data were analyzed for the second year of the study. Data tables were prepared and information was presented to East High School mathematics teachers and the district mathematics coordinator in November 2005. A written report was not prepared. A third set of achievement tests were prepared to administer to students enrolled in the third years of high school mathematics. This set of achievement included a pretest and three forms of a posttest. A matrix-sampling procedure was used for administering the posttests. In addition, teachers were asked to indicate the items on the posttest that their students had the opportunity to learn during the year in order to produce a "fair test" to compare the different curricula. Teachers also responded to a questionnaire reporting on their instructional practices during the year. The third year of data collection of a three year comparative study of mathematics curricula at Madison East High School was completed June 2006.

During the spring and summer of 2006, MMSD conducted a complete revision of their Student Information System. This has entailed a complete overhaul of the student transactional data, revised student record data codes, and tedious uploading of 20 years of data from a variety of individual databases. We have placed a request to the district for the 2005-2006 Wisconsin Knowledge and Concept Examination test scores for individual students and demographic information. These data are necessary to prepare the final report on the three-year study. As of August 2006, the district has indicated that achievement test data are being entered into a database. These data will be provided to SCALE and analyzed in late September and October. The final report will be prepared by the end of October.

The East High School, Madison, Wisconsin, mathematics curricula study was designed to provide information to teachers and district staff to make a decision among three types of curricula—a traditional curriculum, a moderate reform curriculum, and an integrated

reform curriculum. In year 1 of the study, 2004-2005, students enrolled in the traditional algebra course performed significantly higher on the fall pretest than the other students enrolled in the other two curricula. On the spring posttest residual scores given the pretest scores, students taking the moderate reform curriculum and the integrated reform curriculum had above average gains in performance over the 2003-2004 school year.

In year 2 of the study, 2004-2005, this finding was replicated by a second group of students enrolled in the first year of high school mathematics courses. The students enrolled in the moderate reform curriculum had a higher residual score than students enrolled in the other two curricula. Unlike the first year of the study, the integrated reform curriculum students had a lower residual gain score than did the students who took the traditional mathematics curriculum.

The students in the first year of the study were tested again in their second year of mathematics when they took geometry in the 2004-2005 school year. In geometry, students at East High School had the option of taking accelerated geometry course (mainly entering grade 9 students took this course), the geometry course of the moderate reform curriculum, and the second year of the integrated reform curriculum. As expected the accelerated geometry students out performed students in both of the other curriculum on residual gain scores. Students in the moderate curriculum course continued to out perform students in the integrated reform course. After the second year of a three-year study, students in the moderate reform curriculum have had higher gain scores. We are still analyzing data by different demographic groups, by opportunity to learn, and attitudes. There appears to be some differences among the curricula in the content covered as indicated by teachers selecting which items are fair items for their students on the posttests.

SCALE Quality Indicator System - Norman Webb

The focus of the SCALE Quality Indicator System (SQIS) for 2005-2006 has been to acquire achievement data for the current year from the four school districts, design and develop a data warehouse system that can be used to hold and process district data, and to produce reports on achievement data over a period of time from prior to SCALE and with SCALE.

More specifically, individual record data through 2005 was acquired from all four SCALE districts. These data include: student enrollment and course outcome data, teacher certification and assignment data, and student achievement data. The acquisition of data from the school districts is a non-trivial activity that requires a number of steps including approval of data release agreements, the identification and processing of the requested data files from the district staff, encryption of data by the district, entering the data into the SCALE warehouse, checking and verifying the accuracy of the data received, contacting the districts about any discrepancies found, and making the necessary changes to the data received.

The mean achieve scores, as measured by a state assessment, by three of the four SCALE districts—MMSD, DPS, and LAUSD—have varied little from prior to 2003 (before SCALE) and after 2003 (with SCALE). DPS grades 5, 6, and 10 mean mathematics scores had a small increase for 2004 and 2005. DPS grade 11 science scores for 2003, 2004, and 2005 are below the mean scores for 2002 and before. MMSD mean mathematics scores for grades 4 and 8 had a drop in 2004, but then increased to above 2003 levels in 2005 after a general decline in mean scores from 2000 to 2003. MMSD grade 10 scores in mathematics have steadily declined since 2000. In science, the MMSD grade 4 mean scores have had a steady decline from 2000 to 2005, have been relatively flat for grade 10, and have declined for grade 8 until 2005. The mean achievement score for grade 8 science increased in 2005 after a steady declined from 2000. It should be noted that a large amount of SCALE efforts in MMSD has been directed towards middle grades science. LAUSD mathematics mean scale scores for grades 4, 8, and 10 in general have declined for 2003 and 2004 compared to 2002 and before. Trend scores for LAUSD are difficult to interpret because a different test was administered in 2003 (CAT6). The science data is very lean.

For MMSD we have been able to compute the effect size in district mean scores compared to the state mean scores without MMSD. This was done to provide some way of comparing the change in scores to another group on the same tests. For grade 4 mathematics, MMSD out performed the state with an effect size greater than 0.1 in 2003 and 2005. These spikes were different from a general trend of an effect size below 0.05. At grade 8, the effect sizes were greater (nearly 0.2) and also were larger for 2003 and 2005, but not as dramatic as for grade 4 because effect sizes for 2000 through 2002 were higher. The effect size for grade 10 mathematics has steadily declined from nearly 0.4 in 2000 to about 0.05 in 2005. For grade 4 science, the effect size also was higher for 2003 and 2005 (but less than 0.05) than for the other years where the effect sizes were all negative. At grade 8, the effect size for science was the highest for 2003 (nearly 0.1) but were slightly positive for the other years. The grade 10 science effect sizes have steadily

declined from over 0.15 in 2001 to nearly zero in 2005. Thus, there are slight signs of improvement in 2003 and 2005 for grades 4 and 8 mathematics and grade 4 science. More data are needed to determine if these spikes for 2003 and 2005 are a trend or anomalies.

During the year, the SQIS implementation team developed a data warehouse in which to store the large quantity of data received. The warehouse is useful in connecting data types, building analytic data files, calculating metrics, and meeting reporting requirements. This process has required a number of iterations of trying to produce different data reports, verifying the accuracy of the data, and making modification to the data processing procedures. As of July 2006 most of the bugs in the data warehouse have been resolved and the system is operational.

The SQIS implementation team has developed SQL code and logic models to query native data structures from each SCALE district. These queries have been used to calculate NSF MSP-MIS metrics (disaggregated by school, grade, content area, and NCLB categories):

- Teacher counts
- Student enrollment
- Student proficiency status counts
- Enrollment and completion of Math and Science course types (e.g., Algebra 1, Geometry, etc…)

One benefit of this work is the identification of areas in which districts have either relatively high or low capacity. For example, three of the four districts appear to have significant challenges in data quality, particularly related to low reliability between enrollment systems files and human resources files. All districts appear to have some problems tracking student enrollment and teacher assignments for a subset of schools (e.g., some schools appear to have no math or science teachers).

The SQIS implementation team developed Visual Basic for Application (VBA) scripts to automate visual representation of three types of indicators:

- The distribution of scale scores for math and science criterion referenced exams, including means, upper and lower quartiles, 5th and 95th percentiles
- Comparison between NCLB student groups across time, grades, and content area using mean scale score
- Comparison between district mean scale score against state distributions.

These VBA scripts will be used to update graphs as future data are received from districts and will significantly shorten future production cycles.

The SQIS implementation team has used case studies from other research teams to identify SCALE interventions and the likely boundaries of those interventions. Staff are currently developing survey methodology designed for soliciting more recent data about SCALE interventions and the degree to which they are perceived to impact teachers and students.

The SQIS implementation team has acquired individual level data on teacher participation in SCALE related professional development in MMSD. These data span all grade levels for both math and science content areas and has served as a model for how

professional development data may be used to inform the degree to which schools are differentially impacted by SCALE interventions.

Data release agreements for 2006 data are in the process of being sent out in August 2006. A number of data tables have been produced for all four districts. Longitudinal data tables on district achievement have been produced that compare the mean district achievement scores to the state mean achievement scores without the district. Effect sizes have been computed. Reports will be prepared in September 2006.

Over the next year, 2006-2007, the SQIS team will acquire data from the SCALE districts and generate reports on student achievement disaggregated by a number of demographic groupings.

Case Studies of Institutions of Higher Education - Susan Millar

The Institute of Higher Education (IHE) Case Studies line of work for the SCALE RET is comprised of four studies: 1) Case Study of California State University, Dominguez Hills; 2) Case Study of California State University, Northridge; 3) Case Study of University of Wisconsin-Madison; and, 4) Cross-Case Analysis of the three IHEs case studies. The IHE Case Study line of work was initiated in spring of 2005.

The purpose of the case studies is to assess any changes occurring in pre- and in-service training for K-12 math and science teachers, and if the changes can be attributed to SCALE activities. Since these programs are context-specific and complex systems of action, the qualitative case study design was selected. The research design for this line of work is a multi-case design, where research methods are replicated at different sites in order to assess, explore, and describe the context in which SCALE activities take place at IHEs. Each study draws upon in-depth interviews with key faculty and administrators, analysis of documents and reports, and limited observations of meetings and seminars.

Each case study will be comprised of two phases of data collection, analysis, and reporting: a descriptive phase and an exploratory phase. The descriptive phase will collect background material and explore broadly defined topics related to SCALE and pre- and in-service training programs. The exploratory phase will build upon findings from the descriptive phase, focus on emerging themes and topics for further exploration, and assess the ultimate impacts of SCALE on the pre- and in-service programs of each IHE. For each of the IHEs, a preliminary and a final case study report will be developed. A final, cross-case analysis will be conducted at the end of 2007, upon completion of the three final case studies.

During Year 4, the following work was accomplished:

1. Literature review
Linda Scholl began and Matthew Hora continued a review of literature relevant to interactions between teacher education and STEM faculty in IHEs, and to IHE/K-12 partnerships. Hora continues to develop his knowledge of this literature as he proceeds with the IHE case studies.

2. California State University, Dominguez Hills preliminary case study
Interviews with 22 CSUDH administrators and faculty were conducted by Scholl, with the assistance of a graduate student, during fall of 2005. A draft of this preliminary case study was completed in March 2006. During April, Susan Millar obtained approval of the report from all individuals at CSUDH whose input might be recognized, finalized the study, and distributed it to the SCALE RET and key stakeholders at CSUDH. S. Millar presented findings from this case study at the 2006 AERA meeting in San Francisco, CA. Matthew Hora will conduct a second round of interviews at CSUDH in the late fall of 2006, and will prepare the final Case Study by May 2007.

Linda Scholl, Susan Millar, and Latish Owusu-Yeboa wrote the case study report entitled, *Organizational Change in an Institution of Higher Education: Improving K-20*

Math and Science Education through a University-School Partnership. The report includes an initial description and analysis of institutional change efforts underway at one of SCALE's IHE partners, CSUDH. The study found that the organizational climate was perfectly suited for a change effort such as SCALE, due to a predisposition for CSUDH personnel to view K-12 education reform as a critically important endeavor, an unusually large number of faculty retirements, and a campus-wide change effort aimed at making the entire university more student-centered and collaborative. In addition, leaders positioned at various points within CSUDH (provost, deans, chairs, center leaders) and at various points outside of CSUDH (in LAUSD, at UW-Madison) held a largely shared vision for what should be accomplished, and were pursuing diverse, high-leverage, and complementary change strategies, including new funding, and new tools, practices, and policies. The SCALE and QED efforts run parallel to the broader, internally-initiated change process underway across the CSUDH campus that is aimed at improving undergraduate education and linking the services of the university with the needs of the surrounding local communities.

The study finds that one of the most important impacts that the QED and SCALE projects had through fall 2005 at CSUDH was the increased level of trust and collaboration between faculty in the College of Education and the science and math departments. This represents a significant shift from past history when faculty in the College of Education and the math and science departments were largely estranged from each other. Interviewees identified this improving relationship as a product of a number of the SCALE and QED initiatives, including the professional development sessions for STEM faculty run by a faculty member from education, and joint participation in the development of science immersion units for LAUSD. This improving relationship also was supported by broader CSUDH changes in policies and practices intended to create a student-centered campus.

3. California State University, Northridge

Matthew Hora conducted 23 interviews with CSUN administrators and faculty in July and August of 2006. Completion of the preliminary case study is estimated to be October of 2006.

4. University of Wisconsin-Madison

Linda Scholl and Matthew Hora conducted interviews with 21 UW-Madison administrators and faculty in the winter of 2005-2006 and in May and early summer of 2006. This preliminary case study will be completed by mid-August 2006. Hora submitted a proposal to AERA 2007 based on findings of the preliminary UW-Madison case study.

Preliminary findings indicate that SCALE is making progress in each of these areas. Through the *Math Masters* and *Immersion Unit* professional development programs for K-12 math and science teachers, SCALE is engaging STEM faculty in learning and modeling inquiry-based pedagogy, which is influencing the faculty's conception of their own teaching and of K-12 issues. Through the co-construction of professional development materials and the co-facilitation of the actual sessions, SCALE is introducing a new, more collaborative and mutually beneficial partnership between UW-Madison and the Madison Metropolitan School District (MMSD). Finally, SCALE is also

leading inter-departmental efforts to revise the pre-service math and science curriculum for elementary and middle school teacher candidates.

However, the core SCALE strategies face significant barriers at UW-Madison due to an organizational climate that favors research over teaching and service, structural constraints for individual faculty, and pervasive tensions between STEM and education faculty. The primary field of practice that influences SCALE operations at UW-Madison is that of a "Research-One" university, a label that permeates the entire university by reinforcing the importance and to some, the superiority, of research activities in STEM departments. Interviewees also articulated differences between the STEM and education fields, each of which have unique methods, practices, and characteristics that shape the beliefs and behaviors of faculty members over time, as key fields in which they operate. These differences shape individuals' views on teacher education and K-12 issues, and include, for example, opinions on which department is ultimately responsible for educating future teachers, the required coursework for teacher candidates, the appropriateness of IHE faculty participating in K-12 issues, and the role of content based pedagogy.

Despite these barriers, certain aspects of the field of higher education in general, and of UW-Madison in particular, provide faculty with the ability to participate in initiatives such as SCALE. Academic life at UW-Madison is characterized by faculty autonomy. For example, many STEM faculty are sufficiently interested in participating in the teacher education program based on personal experiences with K-12 that they participate significantly in teacher education issues. They do this with impunity as long as they maintain their demanding academic workloads. Additionally, faculty who have substantial social or economic capital often enjoy a degree of freedom and status that allows them to participate in more controversial efforts that attempt to alter the constraints of the organizational context. In most cases these faculty are tenured, as tenure is an extremely strong limiting factor of the academic field. In addition, they usually have relatively high status within their department and college, extensive professional networks, or large amounts of external funding. Several respondents also emphasized the value of avoiding barriers by fostering collaborations in a "neutral space" such as interdisciplinary research centers, which are common at UW-Madison, and where institutional constraints and disciplinary disagreements can be minimized. Phase 2 of the UW-Madison case study will further explore these key themes and issues, and focus on more detailed questions that emerged throughout Phase 1.

Other Work

1. System-Wide Change: An Experimental Study of Teacher Development and Student Achievement in Elementary Science – Adam Gamoran

In addition to the SCALE supported research and evaluation activities described here, a major new line of work was initiated with separate NSF Teacher Professional Continuum funding. Adam Gamoran, Director of the Wisconsin Center for Education Research at UW-Madison, has begun a large experimental investigation of the implementation and effects of elementary school science immersion units. The design involves random assignment of 40 schools to treatment and 40 schools to control. Teacher training began in the summer of 2006. The focus of this major new initiative is to document the effects on student achievement of elementary schools SCALE science immersion units.

The System-Wide Change Experiment is an effort to test the effects of teacher development for immersion unit instruction on student science achievement. Prior evaluation in Los Angeles has indicated that whereas simply distributing immersion curricula did not change teacher practice, distributing the curriculum in the context of an intensive summer institute, with follow-up, changes teacher knowledge and teaching. The institute and follow-up approach to teacher development is expensive and labor-intensive, and the cost of providing it to all 2000 teachers in each grade level in Los Angeles is prohibitive. Consequently the System-Wide Change Experiment was designed to test a model in which 1-2 teachers from each school at each grade (focusing on grades 4-5) will participate in the teacher development activities. Over a two-year period, these colleagues are expected to bring the curriculum and activities to their same-grade colleagues. If this model is effective, resources are available to implement it on a district-wide basis.

In each of the 8 local districts of LAUSD, local superintendents nominated about 20 schools to participate in the study. From the nominated schools, 10 in each local district were randomly selected for the study, from which 5 were randomly assigned to "treatment" and 5 to control groups. Grade 4 teachers from the treatment schools were invited to the institutes during summer 2006; grade 5 teachers will be invited in summer 2007. Outcomes include student achievement on the state standardized test in grade 5, and on periodic assessments in grades 4 and 5. Supplementary studies will observe classroom instruction in treatment schools; test the hypothesis that the effects of the treatment are greater in schools with greater capacity for change; and provide a district-level case study of the scale-up process.

2. Quality Educator Development: Evaluations prepared by Evaluation and Development Associates, September 2006

With resources available through QED formative evaluations of all 27(13 during summer 2005 and 14 during summer 2006) of the SCALE/QED one-week science institutes and five (3 during summer 2005 and 4 during summer of 2006) three-week mathematics institutes were completed. Important information on the processes employed to introduce the teachers to the science immersion units is being analyzed through this work.

Following the initial one-week institutes conducted during the summer of 2005 the professional development facilitators convened to study what information the initial evaluations have given us. This important review helped to guide the work for the next year.

The reports gather information on the number of years of experience of the participants, their attitudes about mathematics/science, their satisfaction with the institutes, the value of the institute topics to the teachers, the strengths of the Professional Development, and if the specific goals of the institutes are met.

Partnership Response to Evaluator's Report

This is the partnership response to the SCALE Evaluator's Report by Andy Porter and the Goal 5 Research and Evaluation Team (RET), which summarizes SCALE research and evaluation findings. As instructed by NSF, the response speaks to "any areas that the Partnership wishes to clarify further" and provides "an indication as to how the findings of the Evaluation Report will influence the next year's Implementation Plan or the overall Strategic Plan."

As noted last year, many of the SCALE lines of work cross over and integrate the original four goals of the project. This is a sign of the success of the SCALE enterprise—not only are the goals progressing, they are combining and reinforcing each other. This Partnership Response draws from each line of work of the SCALE RET in terms of four questions:

- How does the partnership understand and interpret RET findings?
- What adjustments in partnership strategy are appropriate?
- What should be emphasized in future RET research?
- How will the RET teams address the common issue of sustainability?

Building a Partnership

The Building a Partnership (BP) team characterized SCALE working groups in terms of the degree to which they use either "exploitation" or "exploration" approaches to addressing challenges. The Partnership finds this concept a useful tool for analyzing and understanding its different lines of work. For example, now that we are using this concept, we note that the IFL, in its work at the district leadership level, uses an approach that "explores" how to guide and advise district efforts to "exploit" models known to be effective in the four dimensions of the instructional system (Goal 1). By contrast, we note that in pursuing the curricular dimension, the IFL advised LAUSD to choose Cognitive Tutor as an already developed tool to exploit for their algebra readiness initiative. In addition, we note that IFL provided professional development by "exploiting" its technically mature tool of Disciplinary Literacy.

By contrast, we note that it is useful to understand the LAUSD/CSU/SCALE/QED working group that developed and now is implementing science immersion as an example of a group that began by using highly explorative approaches to a very emergent challenge. During 2003-04 and 2004-05, this complex working group literally invented, through co-construction, new curriculum resource and new approaches to professional development, while making improvements based on feedback as these resources were tested, and while also developing new social and administrative networks across the LA basin that will support implementation over time in both LAUSD and the CSU institutions. During 2005-06, this working group has moved toward a combination of exploration and exploitation as they ascertain, based on data, that their resources and processes could be considered technically mature "best practice." In short, we believe that this concept presented by the BP team captures important aspects of SCALE working group. We expect many more BP findings in the near future, and look forward using these findings to further refine our own understandings of leadership and social processes that are at the heart of an effective K-20 partnership.

District Case Studies

The District Case Studies team observed that SCALE has had a significant influence on district policies of instructional guidance and distributed leadership at the top of the system. The SCALE Partnership is pleased with the alignment of the SCALE partnership with district policies and leadership frameworks. Thanks to the case study team, the Partnership is more keenly aware of the need to fully integrate this influence at all levels of the district, including schools and classrooms. This is the only way that SCALE can hope to have a sustainable leadership influence and a coherent use of all district, sub-district, school, and classroom resources in the service of the SCALE goals.

At the SCALE Middle School Math Forum, the District Case Studies team observed that districts and experts converged on a central goal to improve the curriculum knowledge and skills of educators and a set of designs (tools) for achieving this goal. This goal is central to the SCALE theory of action for alignment of professional development in mathematics with the actual curriculum being taught. This approach is based on what research and experience has demonstrated to be effective professional development by providing teachers with both the content knowledge and pedagogical skills (aligned with their current curricula) that adds depth and breadth to their instruction.

The District Case Studies also observed that formative interim assessments in PPSD show promise, but alignment of the assessments to the instructional guidance system and integration at the high school level still pose challenges. PPSD is aware of these problems and has reached out to its SCALE partners in helping to address the challenges it faces in resolving these difficulties. In the fall of 2006 several trips to Providence are planned by teams from the larger SCALE Partnership to address these and other challenges. The broader SCALE Partnership also hopes to benefit from some of the exciting work being done in PPSD in partnership with the East Bay Educational Collaborative.

In LAUSD, the District Case Studies team described differences in breadth and depth of the science professional development and delivery models of immersion at the elementary and secondary levels. The middle school 5-day institute model emphasizes co-development of the immersion unit and co-facilitation of the professional development that engages teachers at an in-depth level of science inquiry. In response to the LAUSD's desire to go to scale district wide with the fourth grade Rot it Right Unit, the elementary model emphasizes a less intensive training approach to a greater number of teachers with a 2-day approach in addition to some 5-day institutes. The current work of the case studies team has provided valuable formative feedback for revisions to the units, and information for improving the institute training and facilitation/leadership sessions. Both the middle school and elementary models will continue to be the subject of research in Year 5 (by the RET district case study team and through Gamoran's TPC study.) The Partnership is keenly interested in what these studies learn about the different approaches to unit development, professional development and delivery, and implementation of immersion in the classroom.

Targeted Studies

The various studies of immersion have also been useful in providing formative feedback to developers, facilitators and our partners as they design and implement the Immersion Units. These studies point out the fact that the Immersion Units are far more than just curricula materials, they encompass the full spectrum of teaching and learning in science programs. The Partnership is pleased with all the science immersion work in LAUSD, MMSD and Pittsburgh to date. This is perhaps the most distinctive new contribution that SCALE has made, and we plan to carefully expand this work, and continue to evaluate the implementation and outcomes, to reach even more of the Partnership's teachers and students.

For example, the Partnership takes very seriously the feedback on the reality of SCALE science immersion implementation in LAUSD. The PERB study focused on the Grade 7 Variation and Natural Selection immersion unit. The report provided process and program feedback that proved useful in revising the SCALE/QED professional development work and follow-up. While the report noted minimal transfer to classroom instruction and practice, the identification of hurdles to classroom implementation was very useful for both the district and for SCALE in planning the needed administrative supports, leadership training, and follow-up for improving implementation this school year. We note that this past summer, the institute enrollment increased and believe that the implementation success rate will increase as well, as the Partnership continues to identify, analyze, and eliminate hurdles.

SCALE Quality Indicator System (SQIS)

The Partnership considers the SQIS work very important for several reasons. The primary reason is that it is an opportunity to measure SCALE impact with the "gold standard," i.e. the impact of SCALE interventions on student achievement. Although casual attribution always will be a problem, the SQIS team is striving to collect meaningful, comprehensive, accessible data that will allow the possibility of analyzing SCALE-related outcomes for students. In addition, quantitative analysis, especially when combined with our qualitative work, will increase the Partnership's ability to improve the implementation of its theory of action.

The SCALE Quality Indicator System team also notes that the mean state achievement scores in three of the four SCALE districts have varied little since the beginning of the SCALE MSP. Given that our work in the first three years of SCALE was largely in development and aimed at educators, it is not surprising that we see little influence of SCALE interventions on student achievement. We anticipate that as SCALE interventions transfer into the classroom, SCALE indicators along with other RET metrics and tools will begin to measure impact on student performance.

IHE Case Studies

The IHE Case Studies team, which began its work in February 2005, produced its preliminary case study of CSUDH in late spring 2005, and is just submitting to us its

preliminary case study of UW-Madison in early October (included with our Year 4 deliverables).

In its preliminary case study of CSUDH, we note the finding that the increased level of trust and collaboration between CSUDH faculty in the College of Education and the science and math departments is indeed one of the most important impacts that the Quality Educator Development (QED) and SCALE projects have had through fall 2005. The Partnership believes that this intra-organizational trust and collaboration is a key component the significant success of SCALE science immersion in the Los Angeles area. This case study is one of the "barometers" that the SCALE Leadership is monitoring carefully, since these new connections are being made and must be carefully nurtured.

We anticipate that the preliminary case study of UW-Madison will prove to be similarly instructive. In particular, we look forward to the IHE Case Study findings on the impact of the Math Masters program on the UW-Madison Math Department, as well as the impact of the UW-Madison's nascent efforts to redesign the curriculum for middle school mathematics teacher certification, and middle school science teacher certification.

Conclusion

Many of the SCALE initiatives are beginning to show promise, leading toward significant institutionalization in both the Partnership districts and IHEs. These accomplishments, and their strengths and weaknesses, have become more visible through the work of the RET. The formative and intermediate feedback the Partnership receives from the RET work enable the SCALE work to grow and simultaneously improve. It is critical that the RET has made the reports accessible to our K-12 district partner as well as our IHE partners. The accessibility of this work to all partners enables direct use for improvement and learning. It is encouraging that the partner districts continue to show increasing reliance on the RET studies and interactive feedback to inform their policies and implementation in math and science.

Since the work of SCALE has continued to expand, we have augmented our evaluation and research resources by engaging other funding sources, such as the QED grant of CSUDH and LAUSD, the Wisconsin MSP Title IIb Math Masters grant with MMSD, and the Teacher Professional Continuum grant of UW-Madison. Susan Tucker, the external evaluator for the QED grant has conducted evaluation of both the science immersion institutes and the math immersion institutes that are jointly run by SCALE and QED. These evaluation reports provide valuable feedback on the SCALE immersion models and professional development sessions. The Math Master's evaluation in Madison has served the reporting needs of both the Title II grantees and SCALE. Moreover, the feedback has been important for STEM faculty and the district, in transforming their approach to in-service and pre-service instruction at all levels of the educational system. Finally, the research work of Adam Gamoran on the TPC grant will provide a rigorous analysis of immersion impact on elementary schools, teachers, and students. These additional studies, taken together with the RET lines of work; promise to provide evidence of the success and lessons for generalization, of the SCALE theory of action, interventions, and partnership. The Partnership looks forward to the continued

contributions of the Research and Evaluation Team to the SCALE Partnership, and also to the national audience that also will benefit from this work.

Publications and Reports: Citations and Abstracts

(June 2005-August 2006)

Building a Partnership

Clifford, M. & Millar, S. (2005). Mapping a partnership: Methodological considerations in describing organization structure and participation. Paper presented at the MSP Summit. September 2005. Minneapolis.

This paper, presented for discussion to MSP evaluators, describes some of the methodological issues associated measuring partnership participation and representing partnership as an organization. Specifically, we discuss how partnerships may officially appear hierarchical and inclusive. Citing problems with organizational charts, we show how partnerships do not necessarily function hierarchically and do not involve high levels of participation by all. We then discuss our attempt to address these representational issues through "organizational mapping" and criteria for "active participation."

Clifford, M. & Millar, S. (2005). What makes for a good partnership? Ad hoc report in response to request from Dan Lauffer, to support an Immersion Team Meeting, November 2005.

This ad hoc report draws upon our and others' research on K12-IHE partnerships. The ad hoc report focuses on key elements of partnership, such as shared goals and trust, and the roles partnership agents (e.g. boundary crossers and upper administrators of partnering organizations) play in the success of partnerships. The ad hoc report was used during SCALE strategic planning.

Lauffer, D., Baxter, H., Clifford, M., Millar, T. & Krinsky, E. (2006) SCALEd Immersion Reform in K-18 Science Education. Paper presented at the AAAS Conference. February 2006. St. Louis.

This conference paper describes SCALE's theory of action, and its development, for scaled math and science reform in the Los Angeles basin. Any reform effort takes time, and SCALE's theory of action in the Los Angeles area is intended to sustain the trajectory of reform beyond SCALE funding. Three aspects of the theory of action were worth noting, based on available literature. First, the theory of action is that involvement of middle-level administrators in reform is important because, in large school districts, middle-level administrators have considerable influence on district program quality and priorities, and they are strategically positioned to influence others. We call this the "middle-out" approach. Second, the theory of action requires organizational flexibility to leverage and respond to changing political, human, and financial conditions. Third, the theory of action recognizes multiple organizations influence teachers' beliefs and abilities to do reform-oriented teaching, so a regional planning approach is necessary.

Case Studies of Institutes of Higher Education

Hora, M. (2006). A Preliminary Case Study of SCALE Activities at the University of Wisconsin-Madison: Factors influencing change initiatives in STEM undergraduate education, teacher training, and partnerships with K-12 districts. University of Wisconsin-Madison (In preparation)

This report of the SCALE Institutions of Higher Education (IHE) Case Studies line of work provides preliminary findings about SCALE activities at the University of Wisconsin-Madison (UW-Madison). This study focuses on the structural and behavioral dynamics influencing the implementation of the four core SCALE strategies for effecting change in IHEs: 1) improving science, technology, engineering, and mathematics (STEM) undergraduate education; 2) improving collaborations between STEM and education faculty regarding pre-service programs; 3) improving collaborations between IHE faculty and K-12 districts regarding in-service training; and 4)) improving institutional policies and practices at the IHE level that support faculty engaged in pre- and in- service activities. Preliminary findings indicate that SCALE is making progress in each of these areas. Through the *Math Masters* and *Immersion Unit* professional development programs for K-12 math and science teachers, SCALE is engaging STEM faculty in learning and modeling inquiry-based pedagogy, which is influencing the faculty's conception of their own teaching and of K-12 issues. Through the co-construction of professional development materials and the co-facilitation of the actual sessions, SCALE is introducing a new, more collaborative and mutually beneficial partnership between UW-Madison and the Madison Metropolitan School District (MMSD). The emerging partnership between the UW-Madison Math Department and MMSD is resulting in greater faculty attention to K-12 issues and needs, and institutional support for the continuation of the *Math Masters* program from both partner organizations. SCALE is also leading inter-departmental efforts to revise the pre-service math and science curriculum for elementary and middle school teacher candidates.

Scholl, L., Millar, S.B., & Owusu-Yeboa, L. 2006. Organizational Change in an Institution of Higher Education: Improving K-20 Math and Science Education through a University-School Partnership. System-Wide Change for All Learners and Educators. Wisconsin Center for Education Research.

This case study report includes an initial description and analysis of institutional change efforts underway at one of SCALE's IHE partners, California State University, Dominguez Hills (CSUDH). The study found that the organizational climate was perfectly suited for a change effort such as SCALE, due to a predisposition for CSUDH personnel to view K-12 education reform as a critically important endeavor, an unusually large number of faculty retirements, and a campus-wide change effort aimed at making the entire university more student-centered and collaborative. In addition, leaders positioned at various points within CSUDH (provost, deans, chairs, center leaders) and at various points outside of CSUDH (in LAUSD, at UW-Madison) held a largely shared vision for what should be accomplished, and were pursuing diverse, high-leverage, and complementary change strategies, including new funding, and new tools, practices, and policies. The SCALE and Quality Educator Development (QED) efforts run parallel to the broader, internally-initiated change process underway across the CSUDH campus that

is aimed at improving undergraduate education and linking the services of the university with the needs of the surrounding local communities.

District Case Studies

Clifford, M. (2006). *SCALE Panoramic Case Study: Denver Public Schools*. Paper presented at the American Educational Research Association, San Francisco, CA.

This paper reviews how leaders in the Denver Public Schools applied a preexisting theory of reform developed for reading to math and science reform, with SCALE assistance. The strategy was to (a) tighten the coupling between central office and schools/classrooms, thus providing a means for central administrators to become instructional leaders; (b) encourage instructional coherence; and (c) increase student test scores. The system load of simultaneously moving to scale with math, science, and reading/writing reform taxed DPS district systems to the point where discretionary funding and external consultants were necessary to directly or indirectly pitch in to accomplish the work. SCALE university partners, particularly the IFL, were very important to this effort, and SCALE funding continued DPS traditions of soft money support for math and science reform.

Clifford, M., Mast, G. (2005). *SCALE Panoramic Case Study: Denver Public Schools*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

This is a case study of the influence of SCALE partnership participation on the Denver Public School district's (DPS) mathematics and science systemic instructional guidance reform efforts. SCALE received funding from the National Science Foundation to improve math and science teaching and learning in large, urban districts. SCALE's strategy is to work through school district personnel and policy to attain sustainable outcomes by supplying human, financial, and material resources from external agencies to increase instructional leadership capacity of district systems.

Clune, W. H. (2006). *Cross-Site synthesis of case studies of system-wide instructional reform in the four SCALE partner district.* Paper presented at the American Educational Research Association, San Francisco, CA.

The impact of the SCALE partnership on district policies of instructional guidance and distributed leadership at the top of the system has been significant. The districts that joined SCALE had theories of action (or strategic plans) that were a good match for the partnership theory. Indeed, the partnership theory was co-constructed with the districts and in many ways was a distillation of common elements. Excellent access to district leadership provided by the partnership fostered a working relationship and yielded rapid development of key elements of policy and organization. Two tasks are incomplete: successfully pushing guidance and support out to schools and classrooms, and building networks of sustainable leadership and support.

Clune, W. H. (2006). *Brief on district needs and designed solutions: SCALE Middle School Math Forum.* Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

The collaboration of districts and experts converged on a central goal—improving the knowledge and skills of teachers around the curriculum as actually taught—and a set of designs (tools) for achieving this goal. Designs included formal professional development organized around big underlying ideas in mathematics, short curriculum units for students on the fundamental concepts present in every textbook, site-based professional development in which coaching becomes part of professional learning communities, inexpensive fast-turnaround formative assessment, and coaching organized around student work.

Clune, W.H. & White, P.A. (2006). *Interim Assessments in Providence Public Schools: preliminary report and issues for discussion*. Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

The overall conclusion is that the interim assessments in PPSD have important strengths, but unresolved problems threaten long term sustainability of the program. On the plus side, there was broad recognition among K-8 teachers of value or potential value of IAs for improving instruction; and support for the IAs remains strong at the central office level despite high turnover of district staff. But problems serious enough to threaten sustainability do exist. Major work is needed in the high schools on standards, assessments, and leadership. Equally important, the problem of how to find time for remedial instruction must be addressed because confusion about how to modify instruction undermines the central purpose of formative assessments.

Osthoff, E. (2006). *Moving from a policy- to system-design focus to increase rigor and equity in teaching and learning: Mathematics instructional reform in Los Angeles Unified School District during the SCALE initiative*. Paper presented at the American Educational Research Association, San Francisco, CA.

This paper shows how Los Angeles Unified School District moved toward a system-design approach to instructional guidance over a period of time. The thesis of the paper is that much of the SCALE-related action in mathematics during the first half of SCALE can be understood as an attempt by SCALE district and IHE (Institute of Higher Education) leaders to work together to help district actors at all levels shift their emphasis away from what can be called a *policy design* perspective and toward a *system design* problematic. In LAUSD during SCALE system leaders have used opportunities for reflection to focus actors' attention on system *coherence*—the extent to which various dimensions of the instructional guidance system are, at the level of practice, working in harmony toward the same intended outcomes. Increased coherence resulted from attention to all instructional guidance tools and interaction among the tools at every level of the system, including classroom teaching and learning.

Scholl, L. (2006) *SCALE Case Study: Evolution of K-8 Science Instructional Guidance in Madison Metropolitan School District.* Paper presented at the American Educational Research Association, San Francisco, CA.

In this case study, I describe the key policy changes taking place within the district both in terms of the overall district framework and within the domain of K-9 science

education. I analyze the factors that enable and constrain the district's capacity for making sustainable change to support a coherent plan for instructional guidance. Finally, I examine the impact of the SCALE partnership on MMSD's capacity to bring a rigorous science program to scale across the district. This case study is based on analysis of data collected between the Fall of 2003 and Spring 2005. Data include: a) multiple interviews conducted with key district staff, b) district documents obtained from the MMSD website, c) documents provided by the district staff, d) observations of district-led science meetings, and e) quantitative data on student enrollment and student achievement in science obtained from the WINNS system (on-line public-access data reporting system on student performance run by the state Department of Public Education).

White, P. A. (2005). *SCALE Case Study: Providence Public School District. Madison.* WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

This case study of the Providence Public School District (PPSD) identifies the effects on district policy and organization of the System-wide Change for All Learners and Educators (SCALE) partnership, with an emphasis on instructional guidance. NSF's SCALE grant funds a five-year partnership focused on improving mathematics and science teaching and learning at all levels. This case study will describe the specific initiatives undertaken in the district to improve teaching and learning in mathematics and science, the extent to which PPSD's theory of change is consistent with that of SCALE, and the role SCALE is playing in the district's change process. The focus of the case study is on the district as an agent of instructional and systemic change. The primary objective of this report is to provide readers with a 'panoramic view' of the mathematics and science initiatives in the PPSD as they intersect with the SCALE Goal One emphasis on the core teaching system, professional development, monitoring, and assessment. Later SCALE case studies of the SCALE partner districts will provide more in-depth investigations of the initiatives as they play out in schools and classrooms.

White, P. A. (2006). *Developing a framework of disciplinary literacy in mathematics in the Providence Public School District.* Paper presented at the American Educational Research Association, San Francisco, CA.

This paper analyzes the policy design and implementation of disciplinary literacy in mathematics in the Providence Public School District (PPSD) as a framework to increase the instructional capacity of schools and teachers. Respondents emphasized that successful implementation of DL involved modifying the entire system, providing coherence and consistency. For the DL framework to be sustained and more fully implemented in mathematics in the PPSD, it will need to become more holistic. As a respondent stated, "The PPSD has the idea, it has kernels to work from, but it needs the threads to tie it together."

White, P.A. (2005). *SCALE Middle School Math Forum participant evaluation summary.* Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

The 2005 SCALE Middle School Math Forum provided a variety of opportunities for both formal and informal conversation to identify useful solutions to pressing issues of mathematics. A summary of the evaluation responses indicates that the participants valued the Middle School Math Forum and gained new information by attending. In comparing the responses of participants from school districts versus post-secondary institutions, the average rating was similar for both categories of participants; however, more school district respondents gave the highest ratings of 4 (a lot) or 5 (a great deal) to both evaluation questions.

SCALE Quality Indicator System (SQIS)

Watson, J. (2005). Using participation maps in the reporting and evaluation of teacher and student involvement in the SCALE Project. Paper presented at the MSP Evaluation Summit held in Minneapolis, Minnesota, September 2005.

Targeted Studies

Apedoe, X., & Schunn, C.D. (2006). *Investigating the tacit problem-solving strategies of novice designers.* Paper submitted to the American Educational Research Association National Meeting, Chicago, IL.

This study is a first step at exploring the problem solving strategies that students use, and the relationship of these designerly problem-solving strategies to students' learning of science when engaged in design-based curricular activities. The purpose of this study is to investigate the problem-solving strategies of students as they attempted to design a solution to a novel design problem.

King M. B., (2006). *Standards and Scoring Criteria for Student Work on Science Immersion***.** Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

King M. B., Davis, D. & Mast, G. (2006). *Electricity and Magnetism Immersion Unit: Summary of Field Test.* Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

In the spring of 2005, MMSD elementary teachers received training on the SCALE Electricity and Magnetism (E&M) Immersion Unit. Three of these teachers, at two schools, implemented the unit in their classes and participated in a field test of the unit. This discussion is grouped into five main sections: 1. an overview of the data we collected, 2. a summary of the observed lessons, 3. a discussion of what we learned from interviews with each of the three teachers, 4. a review of findings from an analysis of student work on unit tasks, and 5. concluding remarks.

Webb, N. L. (2006). *Madison East High School Curriculum Evaluation Study 2005-2006.* Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

The third year of data collection of a three year comparative study of mathematics curricula at Madison East High School was completed June 2006. The East High School mathematics study is a cooperative study between SCALE and the Madison Metropolitan School District (MMSD). SCALE staff has worked cooperatively with MMSD district staff to conduct a study to produce data that can inform the district and high school staff on the effectiveness of different high school curricula and can be used to make decisions about what curricula should be used. This action research and data decision making intervention also is a SCALE targeted study of student achievement over three years.