

System-wide Change for All Learners and Educators

Evaluator's Report

**Submitted to the National Science Foundation
June 2005**

SCALE Evaluator's Report

SCALE Evaluator's Summary	1
Building a Partnership.....	2
District Case Studies	2
Targeted Studies.....	3
Indicator System	4
Case Studies of Institutions of Higher Education	4
Other Work	5
In Conclusion	5
Partnership Response to Evaluator's Summary	7
ABSTRACTS: SCALE Research and Evaluation Team (RET)	11

SCALE Evaluator's Summary

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

There are five lines of research and evaluation work: (1) building a partnership, directed by Susan Millar; (2) district case studies, directed by Bill Clune; (3) targeted studies, directed by Bruce King; (4) building an indicator system, directed by Norman Webb; and (5) institutions of higher education case studies, directed by Susan Millar (this fifth line of work is just being initiated in Year 3). The research and evaluation effort is directed by Andrew Porter, and the goal manager is Sarah Mason.

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. At the end of 5 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 will produce 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 Targeted Studies will ensure that SCALE-produced tools, if appropriately implemented, have the intended effects. Tools found not to be effective will be revised or eliminated. The IHE case studies will document SCALE influences in local IHEs; 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 Los Angeles area suggests that this disappointing history may have a brighter future. 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.

Building a Partnership

The Building a Partnership work has produced several reports to date. One report, *Mapping the Landscape*, identified 80 working groups in SCALE after 16 months from startup, 80% of which are “emergent” in nature, meaning that they are viewed as relatively “soft” (or informal) in purpose and membership. The report revealed that some individuals in the SCALE project are active members of an amazing number of groups. This finding has led the SCALE partnership to re-assess investments of resources. A surprising finding was that district initiatives are so embedded in on-going district work that they are easily overlooked in assessments of SCALE activity. That is consistent with SCALE’s leveraging strategy, but represents a danger in underestimating what the project is accomplishing.

The Building a Partnership team has four sub-studies that are being pursued. The *SCALE Views Study* asks how expectations for and understanding of SCALE goals, strategies, processes, and outcomes vary across SCALE participants. Through “check and reflect” interviews with SCALE project staff in all partner institutions we are learning about lines of communication and size, purpose and stability of working groups. In the coming year, 25 top-level SCALE leaders and an additional 35 SCALE participants will be interviewed. The *Mapping Study* asks how SCALE is structured and how the structure changes over time to meet demands associated with systemic improvement of mathematics and science teaching and learning. Interviews from the SCALE Views study will serve as data. A third sub-study on *Networks* will look into the social networks that contribute to adoption of reform-oriented mathematics and science curriculum, development of improved mathematics and science professional development policies, and how SCALE fits into these social networks. This study is designed in response to the finding that it is difficult to identify whether, and if so how, SCALE influences on-going district work. For this study, interviews and surveys will be used for data collection. The fourth and final sub-study, the *Working Group Cross Case Study*, asks how and why multi-institutional working groups contribute to leadership capacity for mathematics and science teaching and learning improvement. Here, four case studies of SCALE working groups will be completed, followed by a cross-case analysis.

District Case Studies

This past year saw a redesign of our approach to district case studies. Two complementary types of studies are now being conducted in each partner district. One set consists of *panoramic studies* of the impact of SCALE on district instructional guidance systems. The Madison Metropolitan School District (MMSD) case study is in hand and the Los Angeles Unified School District (LAUSD), Denver, and Providence case studies will be completed in May 2005. The second, *in-depth studies*, seek to document the impact of the instructional guidance systems on schools and teachers. For the coming year there are two in-depth studies planned: one on implementation of immersion units in LAUSD; the other on implementation of the core mathematics program in Denver. The MMSD case study identified considerable influence of SCALE on teacher professional development. The case study also revealed that the positioning of SCALE within the district administrative structure has both positive and negative implications for the influence and reach of the SCALE work in the district.

The instruments of change (SCALE as a change agent) is the most challenging aspect of SCALE for district case study work, the one giving rise to problems of causal attribution. None of the goals is advanced through a distinct program fitting the model of program evaluation. Rather, all of the goals are pursued through co-construction of district policy and organization; and the methods of co-construction are multiple, indirect, context-specific, and subject to rapid change, development, and evolution.

The four SCALE goals entail their own distinctive change instruments, such as strategic planning meetings with district leadership, funding of experimental professional development, development of rubrics for district monitoring functions, development and piloting of immersion units, development and piloting of pre-service education, partnership-wide professional development sessions, and others. SCALE also has become

involved in technical assistance ancillary to its main goals, such as providing tools for analyzing alignment of standards, assessments, and curriculum. Goal 3 recently blossomed through alliances with pre-service teacher training programs in LAUSD and Denver. Goal 4 is more exploratory but is beginning to focus on bringing greater coherence to disparate meanings of equity and the objectives of equity programs. In addition to these distinctively SCALE instruments of change, district leadership and coaches receive training from IFL. While the IFL work began prior to SCALE, the content of the IFL meetings already has incorporated various SCALE innovations, most notably the wholesale incorporation of the immersion units into the training of coaches and IFL-wide refinement of SCALE-developed rubrics.

Targeted Studies

The targeted studies work has produced a number of reports. For example, in Denver middle and high school science teachers were surveyed to determine if their students are engaged in deep conceptually-based science instruction on a continuous basis. All teachers surveyed attended the Denver Public Schools Science Teacher Institute. Workshop participants with less experience reported using more fact-based learning methods and fewer inquiry-based learning methods in their classrooms.

In Madison, a 3-year study is comparing the effectiveness of mathematics curricula. The CORE-PLUS mathematics project materials, the Key Curriculum *Discovering Algebra*, and a more traditional course of Algebra I followed by Geometry are being compared in a quasi experiment. Early findings favor the *Discovering Algebra* and CORE-PLUS materials for producing gains in student achievement. Also in Madison, the *MMSD Science Survey, Summary of Results* presents results from the survey of K-8 teachers of science in MMSD. Teachers' positive responses regarding FOSS increased as the number of FOSS modules/courses they taught increased. Elementary teachers discussed student work in science at a much higher rate than middle school teachers.

Reports on each of the three completed courses of the *Math Masters* project show statistically significant increases in participants' content knowledge as measured by pre- and post-tests. STEM faculty at UW-Madison have provided these courses to middle school mathematics teachers from four southern Wisconsin school districts (including MMSD) throughout the 2004-2005 school year.

In LAUSD, pilot studies of immersion units are on-going. Early results have identified the key importance of providing teachers with effective professional development if the immersion units are to be appropriately implemented in classrooms. Variance in quality of implementation presents a continuing challenge.

For targeted studies, the challenge has shifted from too little to study in Year 1, to too much to study in Year 3. As a result, criteria for deciding upon new targeted evaluation studies have been articulated and are being used to focus the work.

Indicator System

The indicator system has been designed to produce valid and reliable information that can be used to (a) monitor the progress of partner districts in terms of student achievement, instructional quality, and teacher professional learning, (b) monitor SCALE progress on specified benchmarks, and (c) report information to NSF to comply with the management information system. A significant challenge for the work is how with limited resources to build an indicator system that provides the results required by NSF, documents the breadth and depth of SCALE influence needed by the project and provides districts with timely useful information for data-based decision-making.

With district partner cooperation, achievement data files have been created for baseline years and the first school year since SCALE's inception. Little change in student achievement was noted over the four years of data collected to date.

Another key indicator in the system is teacher participation in professional development and in teacher instructional practices. Surveys have been designed to collect (a) teacher background information, (b) teacher participation in professional development, (c) teacher instructional practices, (d) school climate, and (e) activities of school administrators. The challenge here is that most of the needed data is not currently being collected by district partners and what data is being collected varies in definition and format across the districts.

Norman Webb presented the indicator work at the 2005 American Education Research Association Annual Meeting. A report has been produced on rate of change and the percent proficient or above in mathematics or science for the baseline years.

Case Studies of Institutions of Higher Education

This new line of work is in response to increasingly strong connections being built by SCALE between local institutions of higher education (that prepare the majority of new teachers for each of our partner districts) and our partner districts. The work will assess (a) whether positive and sustainable improvements are underway in the ways that STEM and education faculty in the IHEs provide professional learning opportunities in mathematics and science and (b) whether those improvements can be attributed, at least in part, to SCALE initiatives.

In order to assess the depth and breadth of institutional change, a snowball sample—starting with deans, and spreading to include other STEM and education administrators and faculty who are more tangentially involved—will be used to track the breadth and depth of improvement to pre-service, induction, and in-service curriculum and pedagogy in K-20 mathematics and science education. A survey component will gather information on faculty and administrator attitudes and beliefs about teaching and K-12 teachers, participation in activities that involve collaboration cross-campus and with K-12 districts. In addition, the effort will gather and analyze documents such as SCALE-related grant documents and program descriptions from the institutions. A proposal has been written and approved; this work is just beginning.

Other Work

In addition to the above 5 main lines of RET work, there are complementary efforts. For example, Chris Schunn has conducted a careful evaluation of the immersion unit that he and his colleagues developed and tested in the Pittsburgh Public Schools. The results of this work are encouraging. Students of teachers participating in the professional development program and using the curriculum materials of the immersion unit gained significantly more in achievement than students of teachers using the appropriate FOSS kit. Further, there appeared to be a narrowing of the achievement gap for students experiencing the SCALE immersion. Yet another example is the study by Andy Porter and Bob Linn of the LAUSD quarterly assessment program in mathematics for algebra and grades K-7. The report concluded that if quarterly assessments are to be a tool for improving the quality of instruction, they need to be seen by teachers as potentially helpful, not potentially threatening. Because the quarterly assessments are used as an “early warning system,” they are seen as part of the accountability system by teachers. The report concludes that a revision of the current system would likely be insufficient to change teacher perception from burdensome accountability to helpful tool. A new system was recommended and is under consideration by the district. A third complementary effort was the content analysis of state content standards and assessments in mathematics and science and of FOSS kits. The results of these analyses were put on an interactive CD and provided to districts at their request.

In Conclusion

As the work progresses, the RET team 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 1-4. 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, the question is “how soon is too soon.” We want to evaluate the impact of SCALE initiatives, but don’t want to reach conclusions about ultimate impact on tools and procedures that are still evolving. 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? 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.

We anticipate that the work will continue to evolve as SCALE matures. For example, while all RET lines of work already are designed to ascertain changes pertaining to the achievement gap, the team has not yet designed a study that focuses explicitly on progress made by the Goal 4 team. As this team's work gathers momentum, the RET will likely initiate additional evaluation activities to capture the nature and effect of its work.

Partnership Response to Evaluator's Summary

This is the partnership response to the SCALE Evaluator's Summary by Andy Porter and the Goal 5 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 Summary will influence the next year's Implementation Plan or the overall Strategic Plan."

As the SCALE Evaluator's Summary notes:

The four SCALE goals entail their own distinctive change instruments, such as strategic planning meetings with district leadership, funding of experimental professional development, development of rubrics for district monitoring functions, development and piloting of immersion units, development and piloting of pre-service education, partnership-wide professional development sessions, and others. SCALE also has become involved in technical assistance ancillary to its main goals, such as providing tools for analyzing alignment of standards, assessments, and curriculum. Goal 3 recently blossomed through alliances with pre-service teacher training programs in LA and Denver. Goal 4 is more exploratory but is beginning to focus on bringing greater coherence to disparate meanings of equity and the objectives of equity programs. In addition to these distinctively SCALE instruments of change, district leadership and coaches receive training from IFL. While the IFL work began prior to SCALE, the content of the IFL meetings already has incorporated various SCALE innovations, most notably the wholesale incorporation of the immersion units into the training of coaches and IFL-wide refinement of SCALE-developed rubrics.

The Partnership also notes that significant SCALE lines of work cross over and integrate to 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. Although this was the intention from the beginning of the Partnership, it is pleasing to be able to report that this integration is becoming a reality. There is a related paragraph in the SCALE Evaluator's Summary:

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. At the end of 5 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 will produce 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 Targeted Studies will ensure that SCALE-produced tools, if appropriately implemented, have the intended effects. Tools found not to be effective will be revised or eliminated. The IHE case studies will document SCALE influences in local IHEs; 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 Los Angeles area suggests that this disappointing history may have a brighter future. 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.

The Partnership would like to add that the partial integration of some of the lines of work across different SCALE Goals now involves cross-institution design and implementation teams. In other words, we have STEM and education faculty, along with district STEM experts, joining together on teams to pursue SCALE lines of work.

The remainder of this Partnership Response draws from each part of the SCALE RET in terms of three 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?

A concluding section considers how the RET teams are addressing the common issue of sustainability.

Building a Partnership and IHE Case Studies

The Building a Partnership (BP) team found that the partnership is composed of a large number of working groups, most of which have emerged around specific tasks and are understood to be transient in nature. This team also found that a few individuals are members of a very large number of working groups and play many roles.

The partnership interprets these findings as appropriate for the early phase of organizational start-up but also as indicating a need for more emphasis on sustainable change in the "second half" of partnership life. For example, while the Partnership considers individuals' membership in many teams vital to ensuring global coherence and minimizing turbulence, we believe it is necessary for these individuals to only play leadership roles (and not play "worker" or even "collaborator" roles) in these teams in order to reduce overload. This is important because overload is inherently unsustainable for the key leaders, and because it leads to blockages in communication and work flow that may hamper efforts to achieve sustainable system change.

The proposed BP studies appear to be aligned with the emerging partnership priority on achieving sustainable change. The studies ask not "Who is doing what?" but rather, "Who is collaborating, how, and to what end?" That is, attention is moving from a description of working groups at a point in time, to a study of partnership interaction processes that impact district and IHE core functions. This is indicated by a shift in research questions being explored. Examples of the new research interests are: 1) How partnership structure is changing to produce systemic change; 2) How IHEs are developing capacity and coordinating with each other and the districts to meet district needs in a coherent way (the IHE case studies); 3) How SCALE has affected social networks of district and IHE administrators; and 4) How multi-institutional working groups are contributing to leadership capacity.

It should be noted that with the new IHE Study, the partnership also will be investigating a parallel set of related questions for the IHEs themselves and for the regional systems of the SCALE districts and their local IHEs.

District Case Studies

The first of the district "panoramic" case studies (Madison) has been completed, with others to follow shortly. The report on Madison demonstrates the usefulness of a panoramic survey of all the major policies of instructional guidance in a district against the framework of the SCALE theory of action. SCALE influences on district policy sometimes flow from discrete organized activities but sometimes must be inferred from the convergence of district policy around the SCALE model.

Considering the decentralized organization of the district, Madison showed strength in two areas of Goal 1 (curriculum standards and professional development), less strength in the other two dimensions (formative assessments and accountability), and a significant lack of alignment of instructional guidance with its local version of IFL's nested learning communities (in Madison, the school improvement process).

We hope that the higher priority given to sustainability in the second half of the partnership will be reflected in increased attention to two related questions: 1) What design and configuration of policies within and across Goal 1 dimensions results in the greatest leverage over change in instruction and learning (for example, the power of appropriately designed formative assessments of student achievement)?, and 2) How can districts be most effective with the newly identified "dimension 5" of Goal 1 (system feedback, management, and problem solving)? Preliminary examples of dimension 5 include increasing sophistication of district decision-makers about the district theory of action (how to design and implement instructional guidance for maximum effectiveness), and appropriate use of formative evaluation of policies.

Targeted Studies

District Targeted Studies provide evidence of the effectiveness of inquiry-based learning. In Madison, inquiry-based Algebra outperformed the traditional Algebra course. In Pittsburgh, middle school students taking an immersion unit gained more than those taking an equivalent FOSS unit, and the achievement gap narrowed between minority students and other students. Familiarity with inquiry teaching also seems beneficial. A Madison survey showed that teachers who taught more units of FOSS had more positive ratings of the experience.

The partnership looks forward to the first targeted studies of newly implemented immersion units in LAUSD. In these and other studies, information relevant to high fidelity implementation and the sustainability of gains would be especially useful, for example, in understanding how to build effective teacher training and professional development around the new instructional units.

SCALE Quality Indicator System

Large parts of the complex foundations of indicator systems for monitoring progress on a continuing basis have been designed, and preliminary trends have been examined. Analysis of trends should become richer and more meaningful as data on additional

variables becomes available and data points extend beyond the baseline into the years that include partnership activity.

Conclusion

Increased emphasis on sustainability in the second half of the partnership is a common theme in the above. For example, the partnership is focusing its efforts on how it can help build and sustain high-level capacity in the core functions of districts, IHEs, and in the regional partnerships among these institutions. In support of this effort, the Building a Partnership and IHE Case Studies teams are undertaking research designed to help both SCALE implementers and all education reformers understand these partnership efforts. The districts will explore and case studies will investigate how to design and configure instructional guidance that really works. Targeted studies will look at the first pilots of immersion units in LAUSD, with attention to components supporting fidelity of implementation, like professional development. Indicators will generate more data over more years to monitor the bottom

ABSTRACTS: SCALE Research and Evaluation Team (RET)

(June 2004-June 2005)

SCALE Quality Indicator System (SQIS)

Webb, N. L. (2005). *Plan of Action for Implementing the SCALE Quality Indicator System*. Presentation at the Annual Meeting of the American Education Research Association. April 11, 2005, Montreal, Canada.

In 2004 and 2005, *The Plan of Action for Implementing the SCALE Quality Indicator System* was enhanced to incorporate the needs of the Management Information System as required by the National Science Foundation. The action plan includes evaluation questions that were derived from benchmarks in the SCALE strategic plan. Next, the critical variables needed to answer these evaluation questions are identified under six major headings: student achievement, student participation, teacher background, teacher professional learning, instruction, and schools. The paper continues by providing examples of possible indicators. One sample is the rate of change in the percent of students who are judged as proficient on state assessments. A second sample is the scale score of teachers responding to questions about inquiry-based learning. The paper concludes with remaining challenges including how to aggregate data in a meaningful way to report statistics that can monitor change over time while representing the theory of change in SCALE's intervention and the assurance that indicators are valid.

Webb, N. L. (2005). *Plan of Action for Implementing the SCALE Quality Indicator System*.

In 2004 and 2005, *The Plan of Action for Implementing the SCALE Quality Indicator System* was enhanced to incorporate the needs of the Management Information System as required by the National Science Foundation. The action plan includes evaluation questions that were derived from benchmarks in the SCALE strategic plan. Next, the critical variables needed to answer these evaluation questions are identified under six major headings: student achievement, student participation, teacher background, teacher professional learning, instruction, and schools. This section of the plan is followed by the identification of the main sources for the needed information and the list of needed instruments. Existing data from the districts will be used to provide information on student achievement, student participation, and teachers. Surveys will be used to gather information on teacher background, teacher professional learning, instruction, and schools.

Webb, N.L. & Kane, J. (2005) *Report of Year 1 of the Comparison of Three Mathematics Curricula at East High School in Madison, Wisconsin*.

This is a report of the first year of a three-year study comparing the effectiveness of mathematics curricula at one high school in Madison. Beginning with the fall semester of the 2003-2004 school year, grade 9 students who enroll at East High School (EHS) in Madison, Wisconsin, had the option of electing to take the Core-Plus Mathematics

Project (CPMP) materials (*Contemporary Mathematics in Context*), the Key Curriculum *Discovering Algebra* (first addition), or a more traditional course of Algebra I followed by Geometry and courses available to third- and fourth-year mathematics students. A quasi experiment design was used including giving those students enrolled in the 9th grade mathematics courses a pre-test in October and a post-test in June. In addition, information was collected on students' opportunity to learn content being assessed, instructional practices, attitudes, and parents' opinions of their child's progress. Data were disaggregated by gender, race, poverty, special education status, and English Language Learner status. After one year, the scores on the posttest, fair and total, for students from all three curricula were very similar. However, because students enrolled in the CPMP course started below the achievement levels of students enrolled in the other two curriculum and those enrolled in *Discovering Algebra* were below those enrolled in *Algebra I* on the fall pre-test, students taking *Discovering Algebra* and CPMP showed above average gains in performance over the 2003-2004 school year. Female students made significantly higher gains than did male students. Those who were not identified as special education students made significantly higher gains than did those who were in special education. There were no significant differences by ethnicity, free and reduced lunch eligibility, and English Language Learner status. However, African American students gained less than average whereas Hispanic students gained more than average.

Kane, J. (2005, February 22). *Tables with Rate of Change in the Percent Proficient or Above in Mathematics and Science For Baseline Years For Four Districts*. (SCALE Quality Indicator System Working Paper.)

Four tables are presented, one for each of the four SCALE partner school districts. The tables list the percent of students judged proficient in the latest baseline year, the range in the percent proficient over the available baseline years, and the average rate of change in the percent proficient. Data are presented by racial/ethnicity groups for mathematics and science as were available. In general, the rate of change was small. For Denver grade 5 and grade 8 mathematics, most groups had positive rates of change from 0.5 to 4.5 percentage points. In grade 8 science the rates of change ranged from 0.5 to 5.0 across ethnicity. Los Angeles rate of change in the percent proficient was the largest of all four districts. In elementary mathematics by ethnicity the rate ranged from 6 to 14 percentage points; in middle school mathematics 0 to 4 percentage points; in high school mathematics it declined from 0 to -2 percentage points. The range in the rate of change for high school science in Los Angeles varied greatly from -12 (American Indian) to 7 (Pacific Islander) percentage points. The rate of change in the percent proficient students for Madison students did not vary greatly, but did show a small decline in mathematics (-2.8 in grade 4, -.1 in grade 8, and -.8 in grade 10). There was a greater decline in the rate of change in proficiency in science (-5.1 in grade 4, -1.6 in grade 8, and -.8 in grade 10). For Providence, the rate of change has not been computed. Over four years in Providence the percent proficient as determined by the three scales on the New Standards assessment (Concepts, Skills, and Problem Solving) was low range for the total group from 37% proficient in grade 4 skills to a low of 5.4% proficient in grade 8 concepts.

Targeted Studies

King, M.B. & Dillenburg, P. (2004, November 12). MMSD Science Survey, Summary of Results.

This report presents results from the survey of K-8 teachers of science in the Madison Metropolitan School District (MMSD). The purpose of this survey is to provide insights from teachers about the science curriculum, science learning environments, and professional development and support in different schools and in the district as a whole. The survey was reviewed and revised by SCALE researchers, district administrators, teachers, principals, and Madison Teachers, Inc. (the teachers' union). The survey was sent to all elementary and middle school principals in late August 2004. Principals were requested to distribute the survey to teachers and return the completed surveys to SCALE researchers at UW-Madison no later than October 1. Participation in this survey was strictly voluntary. A total of 359 surveys were returned representing 46% of elementary teachers and 68% of middle school teachers. Surveys were completed from 26 elementary schools and 10 middle schools.

King, M.B. & Scholl, L. (2004, November 29). *MMSD Science Survey, Preliminary Analysis*, PowerPoint presentation.

The PowerPoint summarizes results from the survey of K-8 teachers of science in the Madison Metropolitan School District (MMSD). It included both quantitative findings and a summary of open-ended questions. The PowerPoint was presented to the MMSD K-8 Science Scope and Sequence Review Committee. Findings will be used for district decision-making on next steps. The survey can be repeated as one measure of instructional change over time.

King, M.B., Sniff, B., & Dillenburg, P. (2004, September 29). *Summary of Content Knowledge Gains for Math Masters Course 1*.

UW-Madison STEM faculty will provide four courses to middle school mathematics teachers from four southern Wisconsin school districts (including MMSD) throughout 2004-05. The first Math Masters course on statistics and probability was conducted from August 20-25, 2004. Thirty-seven teachers participated. In this report, we present findings on teachers' content knowledge from the recently completed Math Masters course. Overall, results show that teachers benefited from the first Math Masters course on statistics and probability. Close to 90% of all participants showed an increase in content knowledge as measured by pre- and post-tests, and gains on four of the eight test items were statistically significant. Participant reflections on the course, both positive and critical, are also summarized.

King, M.B., Sniff, B., & Dillenburg, P. (2005, January 5). *Summary of Content Knowledge Gains for Math Masters Course 2*.

UW-Madison STEM faculty will provide four courses to middle school mathematics teachers from four southern Wisconsin school districts (including MMSD) throughout 2004-05. The second Math Masters course on algebraic relationships/number operations was conducted from October 26-November 20, 2004. Twenty-five teachers participated. In this report, we present findings on teachers' content knowledge from the recently completed Math Masters course on algebraic relationships/number operations. Overall, results show that teachers benefited from the second Math Masters course. Close to 90% of all participants, and 100% of Madison teachers, showed an increase in content knowledge as measured by pre- and post-tests, and average overall gains were statistically significant. Teachers with bachelor's degrees and teachers with graduate degrees showed no differences in gains in content knowledge. Similarly, years of experience teaching math and teaching *Connected Mathematics* were not associated with differences in gains. Participant reflections on the course, both positive and critical, are also summarized.

King, M.B., (2004, September 27). *SCALE Targeted Studies: Criteria and process for determining studies*.

A document outlining criteria for targeted studies and a process for evaluating possible targeted studies was completed. Targeted studies are motivated by three broad research questions: (1) In what ways do SCALE strategies enhance teacher content and pedagogic knowledge in math and science? (2) To what extent do SCALE strategies influence teachers' curriculum, instruction, and assessment? (3) To what extent do SCALE strategies contribute to high and equitable student learning outcomes and improved student interest in science and math? Targeted studies are conducted when they clearly address one or more of the above research questions. But many SCALE activities that could be studied will do this, so the following criteria must be met: utility, feasibility, and potential & scope. A process for determining whether a targeted study is undertaken is also elaborated.

District Case Studies

Clune, W.H. (2005). *District Case Studies Evaluation Design*. Presentation at the Annual Meeting of the American Education Research Association. April 11, 2005, Montreal, Canada.

This paper discusses the design of the case studies of the effects of the SCALE partnership on district policy and organization. The case study design follows the "theory of action" school of evaluation, which requires a focus on the effects of the real operating design of an intervention. Under the SCALE theory of action, change at the level of schools and teachers depends on multi-dimensional change in district policy and practice with three elements: (a) quality and consistency in five dimensions of instructional guidance; (b) redirection of resources; and (c) improved leverage on instruction. The case studies will evaluate how far the districts got in implementing these goals and their effects at the school level.

All of the SCALE goals are pursued through co-construction of district policy and organization at multiple levels; co-construction operates through self-organized working groups whose influence is various, indirect, and context-specific. Small, multiple forms of collaboration, combined with co-construction of policy with much larger and more influential district partners, presents serious problems of causal attribution. The influence of the policies themselves on instruction and student learning presents another challenge. The case studies adopted two basic strategies for dealing with causal attribution. The bottom-up strategy looks for SCALE-like effects in all major examples of target activities, while the top-down strategy looks for the effects of specific SCALE interventions (co-constructions). The top-down strategy is more likely to discover convincing causal links because it looks in areas of high activity, but it is biased toward finding program effects. The bottom up strategy is less likely to establish causal links but is also less biased toward finding program effects.

An ethnographic method was chosen for the case studies because the implementation of the theory of action is completely defined by the local district context. A common outline for the report was developed. Sources of data varied for each district but generally included (a) interviews with district leaders, district middle management and IFL fellows; (b) observation of meetings involving intended change in practice; (c) district documents; and (d) internal data on SCALE activities held in districts, supplemented by data collected by the BP (Building a Partnership) team.

Scholl, L. (2005). *Scale Case Study: Madison Metropolitan School District*.

This case study of Madison Metropolitan School District (MMSD) describes: a) how the district is organizing itself to make and sustain system-wide change; b) the specific initiatives being undertaken to improve teaching and learning in the areas of mathematics and science; c) the extent to which MMSD's theory of change is consistent with that of SCALE; and d) 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. As such, the primary objective of this report is to provide readers with a 'panoramic' view of the initiatives and activities of MMSD as they intersect with the work of SCALE Goal One.

Building a Partnership

Millar, S.B. & Clifford, M. (November 2004). *Mapping the Landscape: A snapshot of SCALE at 16 months: Formative Feedback Report #3*.

This report provides and analyzes a set 14 organizational maps—complex organizational charts showing the formal and informal relationships among individuals as they work to accomplish tasks—that depict the SCALE MSP as it was as of April 2004. Data for the maps is based primarily on 62 interviews with SCALE participants. Two major analytic themes are that: (1) during its first 16 months (January 2003 through April 2004), SCALE expanded rapidly in terms of the number of people and working groups, and that the division of labor is uneven across goal areas and partner organizations; and (2) SCALE working groups serve primarily as an engine for development. Data supporting these analytic assertions include that during this period: SCALE grew to include 178

active participants (defined as individuals who are aware of the SCALE project as such, and participate in ongoing work that they and at least one other individual from their institution believes is contributing to the achievement of SCALE goals); these participants were organized into 80 working groups, 57 of which were formed between September 2003 and April 2004; and 63 of the 80 working groups were classified as “emergent,” meaning that they are viewed as relatively “soft” (or informal) in purpose and membership. Other key findings are that the professional roles of participants generally were as expected, given the SCALE plans of action articulated in administrative documents, and that the predominating proportion of active participants were from institutions of higher education. Upon categorizing participants’ roles in working groups (worker, manager, director, collaborator), the 178 active participants held 567 roles. Of note, 32% of these roles were in the “collaborator” category. The number of roles per participant was distributed unevenly, with a few individuals playing 10 or more roles. Also of note is that, when the working groups were classified by function (problem identification, development, coordination, dissemination, evaluation and research), the predominating proportion was found to be engaged in the development of products, events, and designs intended to address partnership goals.

Clifford, M. & Millar, S. (2005). *Organizational Mapping as a Tool for Understanding K-20 Partnerships to Improve Mathematics and Science Education*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Organizations in the “instructional change industry” populate the space between statehouses and schoolhouses, and their function is to work with each entity to improve conditions for student learning by conducting research, developing instructional materials, consulting with leaders, providing funding, and training personnel. To implement instructional change policy, government agencies have turned to educational partnerships, particularly those involving K-12 districts and institutions of higher education. In this paper, we aim to describe how and why one K-12 and higher education partnership is structured to improve mathematics and science learning for all students in large, urban school districts. Initially, we sought to collect data via organizational charting, but found organizational charts inappropriate because partnership structures were embedded in participating organizations. We discuss “organizational mapping” as a method for attending to formal, informal, and embedded partnership divisions and sub-divisions. Using organizational mapping, we were able to quantify the types and roles of sub-divisions, participants, and work taken on by the partnership. We found that “working groups,” temporary task forces involving K-20 representatives formed to achieve tasks, dominated the partnership. The majority of participants in partnership working groups hailed from, and were led by, institutions of higher education, and the majority of working groups were located within institutions of higher education. Working groups were tasked to identify problems to be addressed by the partnership, coordinate resources, design materials, and evaluate work.

Additional SCALE Publications

Bohanan, R.E., Niemi, K.J. & Wachtel, L. *Motivating Students to Ask Scientifically Productive Questions*. Published in the Ecological Education column of the Bulletin of the Ecological Society of America, April 2005.

Mehalik, M., Doppelt, Y., Schunn, C. (Draft submitted for publication in AERJ in 2004.) *Middle-School Science Through Design-Based Learning versus Scripted Inquiry: Better Overall Science Concept Learning and Equity Gap Reduction*.

Porter, A., & Linn, R. (2005). *Report on LAUSD quarterly assessments in mathematics*. Submitted to LAUSD.

The authors conducted a review of the LAUSD quarterly assessment program in mathematics for grades K-7 and for algebra. The analysis led to the following recommendations. First, LAUSD should create a new system that would be received by teachers as a helpful tool rather than burdensome accountability. Once a manageable set of standards has been identified for each grade, a test should be developed for each standard. The test should have sufficient validity and reliability to be useful for determining student mastery of the standards. The criterion of mastery should be carefully set, perhaps using a Bookmark procedure. The test should be given and scored by teachers under standardized conditions and supported by appropriate professional development. The teacher should decide when students should be given a test for a particular content standard. Second, there should be an on-going evaluation of the quarterly assessment program. The evaluation should judge the quality of the instruments, their administration, the reporting, the use made of the results, and the effects of the program on instruction and student achievement. The LAUSD might create a small technical advisory panel of three or four outside experts to guide the work.

Porter, A. (in press). Prospects for school reform and closing the achievement gap. In C.A. Dwyer (Ed.), *Measurement and research in the accountability era*. Mahwah, NJ: Lawrence Erlbaum Associates.

The paper describes the achievement gap—how big, how stable over time, how stable over age, and how important—and then considers each of several reforms that have been hypothesized to decrease the achievement gap. The achievement gap continues to be large, whether defined by race/ethnicity or socioeconomic status. The gap is large for all ages and for all academic content areas. Although the gap narrowed during the 1970s and early 80s, since then the gap has been relatively stable. The achievement gap is not the same across academic subjects. It is small for reading than for mathematics or science. It appears to increase less with age for reading than for mathematics. Further, summer loss is greater in mathematics than in reading.

The most promising strategies for reducing the achievement gap appear to share the ability to address inequalities in opportunity to learn. Teacher quality appears to be an especially promising avenue. Currently, Black, Hispanic and low-income students are

less likely to be taught by high-quality teachers than are White and more affluent students. Similarly, removing inequalities in the content students study is an especially promising reform for reducing the achievement gap. One possibility is to eliminate basic and remedial courses in high school, requiring instead that all students take college preparatory courses in core academic subjects.