THE IMPACT OF EARLY POSITIVE RESULTS ON A MATHEMATICS AND SCIENCE PARTNERSHIP: THE EXPERIENCE OF THE INSTITUTE FOR CHEMISTRY LITERACY THROUGH COMPUTATIONAL SCIENCE

K.S. MURRAY and M.A. HENRY
M.A. Henry Consulting, LLC
Co-Lead Evaluators, Institute for Chemistry Literacy through Computational Science
St. Louis, MO 63119
keithsmurray@mahenryconsulting.com; mahenry@mahenryconsulting.com

M.C. HOGREBE
Dept. of Education, Washington University in St. Louis
St. Louis, MO 63130

Abstract

After one year of implementation, the Institute for Chemistry Literacy through Computational Science, an NSF Mathematics and Science Partnership Institute Project led by the University of Illinois at Urbana-Champaign's Department of Chemistry, College of Medicine, and National Center for Supercomputing Applications, experienced statistically significant gains in chemistry content knowledge among students of the rural high school teachers participating in its intensive, year-round professional development course, compared to a control group. The project utilizes a two-cohort, delayed-treatment, random control trial, quasi-experimental research design with the second cohort entering treatment one year following the first. The three-year treatment includes intensive two-week summer institutes, occasional school year workshops and year-round, on-line collaborative lesson development, resource sharing, and expert support. The means of student pre-test scores for Cohort I (n=963) and Cohort II (n=862) teachers were not significantly different. The mean gain (difference between pre-test and post-test scores) after seven months in the classroom for Cohort I was 9.8 percentage points, compared to 6.7 percentage points for Cohort II. This statistically significant difference ($p<.001$) represented an effect size of .25 standard deviation units, and indicated unusually early confirmation of treatment effects. When post-tests were compared, Cohort I students scored significantly higher than Cohort II and supported the gain score differences. The impact of these results on treatment and research plans is discussed, concentrating on the effect of lessening rural teachers' isolation and increasing access to tools to facilitate learning.

Introduction

When to expect outcome data sufficiently robust to assist research design and implementation refinement is a subject of general interest in the treatment of human subjects in education programming. The answer, at least in the authors' experience evaluating mathematics and science partnerships funded through the National Science Foundation or the Department of

The Journal of Mathematics and Science: Collaborative Explorations Volume 11 (2009) 95 - 108

Education and similar projects, has been later rather than sooner. Effects of teacher professional development programs on student achievement are often seen as long range outcomes beyond the three-to-five-year span of grant programs and their research components [1, 2]. The long-term course required to affect teacher performance measurably, with its attendant complexities, is accepted as a reasonable given [3, 4]. Additionally, study designs and research efforts can be constrained by resource availability, variable project staff and participant cooperation, extant data limitations, and the need for evaluative focus on formative and process concerns to ensure fidelity of implementation [5, 6].

Project leaders and evaluative researchers often must make do with the basics—pre-/post-tests framing relatively brief treatment phases, self-reported change in classroom practice, and limited classroom observation—which are perhaps the most commonly applied measures used to investigate achievement effects [7, 8]. However, if a sufficiently rigorous research and evaluation design is in place, if project cooperation is sufficiently supportive of research efforts, and if project activities are implemented with vigor, intensity and fidelity to plan, what may be expected? When can outcome data sufficiently robust to guide future implementation and research activities be developed? Put another way, what is the impact of such early analyses and results if they are available?

This paper addresses the case of the Institute for Chemistry Literacy through Computational Science (ICLCS), a National Science Foundation Mathematics and Science Partnership (MSP) Institute Project led by the University of Illinois at Urbana-Champaign's (UIUC) National Center for Supercomputing Applications (NCSA), School of Medicine, and Department of Chemistry. The ICLCS is a five-year research project investigating the effects of a statewide teacher professional development effort aimed at rural Illinois high school chemistry teachers. Thom Dunning, a professor in the Department of Chemistry and Director of NCSA, is the project's Principal Investigator. In addition to UIUC, other core partners for the project are the AC-Central School District and the Regional Office of Education #38, both rural educational entities in central Illinois.

The project includes the following goals:

- Improved teacher and student content acquisition in the context of present-day research;
- Increased teacher comfort with and use of computational and visualization tools in the classroom;

- Teacher-leadership development in STEM and computational science education; and,
- Related institutional change at the University and among the K-12 educational partners engaged in the project.

The project, funded in 2006, has just entered its second year of treatment for one cohort of teachers and its first year of treatment for the second cohort. This second cohort also serves as the control group for the first cohort. Treatment includes the following components: an intensive two-week summer institute conducted annually for three years for each cohort; ongoing virtual learning community activities through work group assignments, lesson planning, resource sharing, and rapid-response support to teachers' questions; twice annual workshops; provision of tools and technical support to teachers for use in their classrooms; and, individual leadership development planning. Central to project communications and activities is the use of a centralized, on-line system through which almost all ICLCS contacts, assignments, work products, resource information, etc. are shared. Teachers will be followed for two years after the formal treatment course. This article describes the research design and methods used for the project, reports early results, and discusses some of the effects of these early results on the project, evaluation, and research plans and activities.

Methods

Research and evaluation design and implementation for ICLCS is the responsibility of an external team from M.A. Henry Consulting, LLC, a St. Louis-based educational research and evaluation firm. External evaluation is a requirement of NSF Institute MSP projects.

Methods—Recruitment and Ascertainment

Teacher participants in the ICLCS were recruited through a broad-based effort that included information shared with Illinois state and local educational leaders and professional organizations, presented on various listservs, and communicated to more than 300 teachers who had expressed interest in an earlier needs assessment effort. A second focused recruitment aimed at areas of the state underrepresented in the first wave of results. Acceptance criteria included the requirement that teachers were currently teaching high school chemistry in an identified rural school district, an agreement by the principal and district to cooperate with project technical and teacher time requirements, and a personal statement of commitment by the teachers.

Methods—Random Assignment and Research Design

A quasi-experimental, two-cohort research design based on random assignment and delayed control group treatment was developed for the project. Versions of this design had been recommended at a joint Department of Education/National Science Foundation MSP conference as appropriately rigorous within the constraints usual in educational research [9]. Cohort I was to serve as the initial treatment group, with Cohort II serving as the control group with treatment delayed until the following year. As time passed, the second cohort would continue to be used as the control, as it always would be one year behind the first cohort's treatment. The research design is outlined in Table 1.

Once recruited, teachers were listed in order by their district's standardized mathematics scores, and randomly assigned by pairs into one of two cohorts. The first cohort was identified as the first treatment group. In the three cases where more than one teacher had applied from the same district, all teachers in that district were assigned to the same cohort, for purposes of resource sharing and avoiding cross-cohort contamination. Teachers recruited after cohort selection were included in project participation, but excluded from analyses focusing on the core treatment and control groups.

Table 1
ICLCS Research Design: Teacher and Student Chemistry Content Testing

Cohort	Year 1	Year 2	Year 3	Year 4	Year 5
Cohort I	Teacher Identification, Treatment Year 1	Treatment Year 2	Treatment Year 3	Re-test of content for retention	Re-test of content for retention
Cohort II	Teacher Identification, Control for Cohort I	Treatment Year 1, Control Cohort I	Treatment Year 2, Control Cohort I	Treatment Year 3	Re-test of content for retention
	Teacher ACS testing and Student ACS testing, both cohorts, all years				

Methods—Teacher Cohort Characteristics

The initial research cohort contingent totaled 101 teachers. Early pre-treatment attrition, reassignment from the research cohort to non-research cadre participation in the other group because of teacher issues, and subsequent attrition left n=38 for Cohort I and n=39 for Cohort II, for a total of 77 teachers available to participate in the first treatment year's research activities.

Table 2
ICLCS Project Teacher Participants: Research Cohorts and Total Participants

	Initial Research Cohort	Number in Research Cohort	Percent Research Cohort Retained	Total Teachers Recruited (Research Cohorts and Non-Research Participants)	Total Retained	Total Percent Retained
Treatment Group 1 (Treatment in Years 1-3)	51	38	75.5%	51	44	86.3%
Control Group 2 (Treatment in Years 2-4)	50	39	78.0%	69	60	87.0%
Total	101	77	76.2%	120	104	86.7%

Teachers in both cohorts had a broad mix of educational backgrounds and teaching assignments. While all teachers were engaged in chemistry teaching, many also taught physics, biology, general science, and other subjects. Some also had earth science responsibilities or worked with advanced chemistry courses. Eighty-one percent had undergraduate degrees in a science subject, with 36% having general science degrees, 19% having biology or biology education degrees, and 14% having chemistry degrees. Forty-seven percent of the teachers had graduate-level degrees. Of these graduate degrees, 78% were in the sciences or science education, but only 10% were specifically in chemistry or chemistry education.

Methods—Project Activities: Treatment

Treatment for the teachers in Cohort I has been described previously. The project has committed itself to design and implement its curriculum around the stated and demonstrated needs of the teachers, and to integrate computational and visualization tools into their real-world classroom work. The project team does not attempt to dictate new curriculum. Rather, its focus is on assisting teachers in integrating computational tools and content support into their existing diverse chemistry curricula. The Summer Institute represents more than eighty hours of intensive work, with most day schedules including twelve hours of activities. Treatment comprises a combination of resource sharing, content refreshers, leadership workshops, open labs, and small group work engaged in lesson module development. Ample opportunity is given for teachers to address the concerns and challenges they face in their own classrooms. These and other project activities are organized into a graduate-level chemistry course for the participating teachers, which encompasses the academic-year workshops and project engagement with the virtual learning communication system that connects project participants, faculty, and staff throughout the year.

University faculty, drawn from computational chemistry, general chemistry, bioinformatics and computational biology, biochemistry and molecular and integrative physiology, and instructional development areas work closely with the teachers, aided by other staff and graduate and undergraduate students. Additional chemistry and medical school faculty serve as mentors assigned to each of the teacher groups engaged in lesson module development. The project also has provided and has been assisting in installing Personal Interfaces to the Access Grid (PIG's) at teachers' schools, with cameras and headsets provided. Technical limitations at the schools have presented a predictable challenge, but to date, twenty-eight of these PIG's have been installed to enable teachers to communicate with other teachers and participate in real-time technology and content refreshers during the academic year.

Methods—Chemistry Content Analysis Measures and Procedures

To establish chemistry content knowledge baselines, Cohort I teachers completed the American Chemical Society's *General Chemistry Brief Test for the Full-Year Course* at the start of their first Summer Institute [10]. To measure gains, this same test was given at the start of their second Summer Institute for post-test purposes. Analysis of these data are ongoing. Cohort II completed their comparable baseline pre-tests at an informational meeting in Spring 2008, three months prior to the start of their first Summer Institute, and will take post-tests at the start of their second Summer Institute in July 2009.

This teacher testing schedule is used in order to capture the effects of yearlong treatment, rather than merely the brief, intensive work in the Summer Institute. The project asserts that its sustained engagement with teachers during the school year via the on-line system, workshops, and teacher group work will enhance gains in content knowledge, as well as classroom practice and incorporation of computational tools in chemistry lessons. Post-tests, therefore, are timed to capture the effects of a full year's engagement in the project.

Student chemistry concept testing is performed using the American Chemical Society's *High School Chemistry Test* [11]. Pre-tests are to be delivered at the start of the school year to students of both Cohort I and Cohort II teachers, with post-tests delivered by the beginning of April to accommodate timing of Illinois state standardized testing.

American Chemical Society (ACS) tests were selected because of their long established status, broad acceptance, and coverage of appropriate chemistry concepts. Teachers in ICLCS noted that they did not teach all concepts on the ACS tests to their students. However, as chemistry content domains on the tests represent the full spectrum of Illinois chemistry high school standards, measures for all domains were included in testing.

Content tests also are delivered to non-research teacher participants in both groups and pre-tests and post-tests are delivered to their students. Parallel analyses are performed for the full cadres at the same time as research cohort analyses are done. Teachers themselves are not informed whether they are part of the research cohorts, although the circumstances of their entry into the project could inform them of their status.

Student ACS tests are delivered to students by their teachers at their schools. Teachers score their own student tests and report results using individualized student codes. Scores and answer sheets for pre-tests and post-tests are returned upon completion to the research and evaluation team for data entry and quality control checks. The test copies are returned with post-test materials for redistribution at the start of the next year's school year.

Other data are collected from teachers in numerous ways. Surveys track confidence with chemistry content domains, access to and use of technical resources, support networks and teaching workload. Interviews, classroom observations, module plan analysis, and analysis of communications on the on-line system are among the other data collection methods being applied to the project. In the case of this article, however, ACS content tests serve as the item of focus.

Results

Matched pre-/post-tests were returned by fifty-four of the seventy-seven research cohort teachers, for a 70% response rate. By cohort, twenty-nine of thirty-eight Cohort I teachers (76%) and twenty-five of thirty-nine Cohort II teachers (64%) contributed matched pre-/post-tests. The total number of students for whom matched pre-/post-tests were returned was 1,825, of whom 963 were students of Cohort I teachers and 862 were students of Cohort II teachers. The mean number of student tests returned by teachers was thirty-four. The results of these tests provide the first evidence of whether or not ICLCS teacher participants are contributing to an effect in student achievement in chemistry content areas.

The results are reported in Table 3. The mean pre-test scores were not statistically different for students of Cohort I teachers and students of teachers in Cohort II (t = -0.016, df = 1,823, *ns*). The mean pre-test score was 27.4% correct for both Cohort I and Cohort II students. This result supports the comparability of students' chemistry knowledge between the two cohorts and appears to indicate the soundness of the random assignment process used. Student pre-test scores ranged from 0 to 65.0% correct for the Cohort I treatment group and 0 to 67.5% for the Cohort II control group.

Table 3
Pre-Test, Post-Test, and Gain Score Mean Differences:
Research Cohort I versus Research Cohort II

	Group	N	Mean	Standard Deviation	Mean Difference	*t*-test
Pre-Test	Cohort I	963	27.407	8.106	-0.0065	t = -0.016
	Cohort II	862	27.413	9.065		Not significant
Post-Test	Cohort I	963	37.17	12.823	2.996	t = 5.32
	Cohort II	862	34.18	11.219		$p < .001$
Gain Score	Cohort I	963	9.765	13.106	3.042	t = 5.13
	Cohort II	862	6.723	12.098		$p < .001$

Comparison of post-tests between research cohorts and total cadres presents a very different picture. Cohort I students scored significantly higher (t = 5.32, df = 1,821, $p < .001$) on their post-tests than did Cohort II students (mean difference of 2.996), with a Glass's effect size of .27 of a standard deviation. The range of post-test scores was 0 to 80.0% correct for the treatment group and 0 to 78% for the control group. Running this analysis with pre-test scores as a covariate to increase the power of the analysis yielded similar results.

Finally, when pre-test and post-test results are considered in the context of comparative gains, additional differences between the Cohort I treatment group and Cohort II control group are evident. Results for gain score differences are similar to those for post-test differences. Cohort I students scored significantly higher gains on their post-test than did Cohort II students with a mean difference in gains of 3.0419 ($t = 5.132$, $df = 1{,}823$, $p < .001$) and effect size of .25 standard deviation units.

To summarize, given the basis of research cohorts in a quasi-experimental, randomized assignment design and the lack of pre-test differences between the two cohorts, a significant difference can be seen in measures of treatment effects of the ICLCS project on Cohort I teachers over the control group (Cohort II) in terms of content acquisition by their students, as evidenced by differences in ACS chemistry test scores among students.

Discussion—Caveats

The finding of statistically significant greater gains in chemistry content knowledge among students of treatment teachers versus students of control teachers following at most ten months of teacher treatment is unusual. Before considering the impact of these results, discussion of some caveats is useful.

The possible effects of the teacher response rates in returning student tests must be taken into account. As described, an overall 70% response rate was experienced, based on returns of matched pre-tests and post-tests. By cohort, the rates were 76% returns by treatment teachers and 64% by control teachers. It is conceivable that differences in responder and non-responder characteristics among teachers in the two cohorts could have contributed at least some of the apparent differences in gains seen. A lower response rate among control teachers could indicate less motivation generally, as one may expect active participants to respond at higher rates.

Non-responding teachers in both cohorts reported uniformly to evaluators that their reasons for not returning student tests were either confusion about procedures, or workload issues at school and at home that prevented undertaking this extra work. To counter the explanation of confusion, it can be noted that non-responding teachers received no fewer than six reminders from evaluators and project staff during the course of the school year. No obvious difference in non-responding teacher characteristics between cohorts was seen when teaching experience or length of time at their schools was compared. Finally, the actual difference in the number of teachers returning student tests between treatment teachers and control teachers was only four teachers, with twenty-nine treatment and twenty-five control teachers responding.

Despite the possibility of some differential response effect on gains, the non-equivocal strength of the statistically significant differences, the relatively large numbers of student scores included in the analyses in both cohorts, and the lack of a discernible pattern in cohort non-response characteristics indicate to the authors that a positive treatment effect in students at an early stage in the teachers' involvement in ICLCS is evidenced by analysis results.

Discussion—Impact of Early Positive Results on Student Chemistry Content Knowledge

First, the availability of such results is a direct consequence of the quasi-experimental design in place for the project. With a less rigorous design, confidence would be reduced and a greater chance would exist that positive, negative, or inconclusive results could be missed.

Of course, in simplistic terms a persuasive indication of positive student effects from a short-term teacher professional development treatment represents a welcome scenario. The results have served to informally validate both the project plan and the research design developed to investigate it. Based on observation and informal interviews, some stakeholders engaged in the project with a layperson's view of evaluation, acquired a greater understanding of the usefulness of the design. For example, this understanding has reduced requests for cross-cohort mingling in the interest of sharing helpful information more broadly. Such enhanced cooperation is not trivial. In educational research, it can be challenging to communicate about design protocols convincingly so that participants who are unfamiliar with such work do not view these procedural requirements as counterproductive or unnecessarily draconian.

In a similar vein, evaluators were asked by project leaders to share selected results with the treatment group, partly in response to teacher requests for information and partly to see if such information could assist in strengthening cooperation with evaluation procedures. Cohort I treatment teachers were briefly told of student content gains results in a group session during which their own ACS post-tests were delivered. The response was overwhelmingly supportive, with numerous questions posed for the first time about research plan rationales. This response and the stated commitment of several teachers to conform more closely to the research model resonates with literature on the benefits of teachers' active participation in research [12].

When asked in this session what the teachers themselves thought about the gains seen in their teaching, an overwhelming majority agreed that the project was providing them with a network of content support and engagement previously lacking in their work. Furthermore, it was also providing them with tools that they could apply in their existing curricula to better stimulate their students' interest and enhance their classroom delivery.

The limited availability of other teachers or content experts to consult when questions or challenges arise has been documented by the authors among these rural teachers. As discussed, they often are teaching multiple subjects and may be the only high school science teachers in their districts. The capacity of the ICLCS project to meaningfully connect them with other teachers and with content authorities, as represented by university faculty, confirms its relevance to the teachers' actual teaching practices.

Additionally, technical resources and support often are lacking in rural schools and districts. At times, the issue is not just equipment, but installation and troubleshooting. The project has connected teachers not only with various tools, but when challenges arise, the project has provided technical assistance, directly to them or to the technical staff at their school or district. Despite continued difficulties in some schools concerning computer and Internet access, including bandwidth concerns, teachers stated that they felt better equipped for chemistry to engage students more actively and meaningfully. Another potential benefit from sharing results with teachers and thereby increasing their understanding and motivation about the research component is the possibility that response rates could improve when next year's student ACS tests are to be delivered.

The impact of the early positive results on the project implementation team has been to support the curricular choices made for the Summer Institute and other project activities. Allowing much of the chemistry content choices to emerge from teacher needs appears defensible. The project has recognized the difficulty of strongly prescribing curriculum in a treatment setting including teachers from eighty different school districts. Teachers with varying degrees of experience and understanding have shown themselves to be reliable arbiters of what they need to improve their chemistry teaching. The project team also agree from their own perspectives that the reduction of teacher isolation is largely explained by immersion in an ongoing learning community and through the provision of computational, visualization, and other chemistry tools to aid in lesson planning.

An added potential boon anticipated from early student outcomes related to the project is their effect on schools and school districts of the participating teachers. To date, school and district support for the project has been related more to teacher release time for the two workshops and the acceptance of some equipment and technical assistance related to PIG installation. It is expected that demonstrable positive project effects on their students will help lift the collaboration with schools and districts to the next level. This step will be useful as the project attempts to extend its beneficial influence through teacher leadership plan implementation and possible connections of the computational and visualization tools to other subject areas in the sciences.

The student results also have assisted the evaluators in expediting plans for more in-depth multivariate analyses to determine more specifically what causal chains may be at play between ICLCS participation, and teacher and student outcomes. Against the possibility that teachers in the treatment group and their students were in such uniformly dire straits concerning chemistry learning that any treatment was likely to produce an immediate, if short-term, effect, the evaluators note the range in student pre-test scores in both treatment and control groups. Of course, the next series of student test results will provide further indications of the longer-term pattern of student content acquisition post-teacher treatment.

An additional area of interest is in developing long-range plans to adapt the ICLCS model for replication in other circumstances. The availability of NCSA, a first-tier research organization in computational science, and other chemistry resources at UIUC makes for a project plan difficult to apply elsewhere. The indication that intensive and sustained engagement by many senior faculty members and active research scientists in high school teacher professional development can have a positive effect on teachers and their students could seem less noteworthy than the question of how such experience can be adopted elsewhere. The project team already has begun to address this issue, and is planning to refine virtual learning community tools and dissemination of computational and visualization tools usable in diverse educational contexts.

Future Directions

Further analyses contain the following variables of interest: depth and focus of teacher engagement in ICLCS; teacher formal education and degree concentration; teaching experience; school and district demographic characteristics; extent of school and district support—general to chemistry and science curriculum and project specific; content, pedagogy and technical support network changes; intellectual leadership growth; confidence with content domains; and, observed classroom practice. The interrelationships of these variables and their role in student content gains and other student outcomes will be investigated. Another possible area to consider is the project effect on high school student interest in and pursuit of college enrollment, particularly with chemistry or other science majors in mind. While many of these analyses have been planned since the project's inception, the early and unexpected student gains seen have helped frame these analyses and contextualize them.

Case study methods will be applied to an in-depth study of nine teachers in Cohort I in order to better understand the work lives of teachers. Selected because of their middle range of teaching experience and their prior full participation in the project and its research components, these teachers will be visited and observed in the field for three days at a time by the research team. These observational and related teacher interview data will be augmented by interviews with school and district staff, including principals, superintendents, other teachers, and technical support staff members.

Conclusion

Finally, the availability of early positive student content gain data has assisted in further coalescing the partnerships contained within the ICLCS project. While partnership is inherent, as well as explicit, in the NSF Mathematics and Science Partnership program, the extension of partnership models for more in-depth exploration into the inter-organizational and interpersonal workings of implementation and research is facilitated by the first objective evidence of project success. Reaching students and being able to demonstrate this often reside at two different points on the educational research map. To have an early indication of project efficacy in affecting student content knowledge is both gratifying and a challenge to the project for continued rigorous and engaging work. Inclusion of outcome data analyses as early as possible in the implementation phase during research design development, regardless of the outcomes that emerge, helps ensure both the means to confirm efficacy and to indicate refinements called for in order to achieve the project's stated objectives.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 0634423.

References

[1] L. Desimone, A.C. Porter, M. Garet, K.S. Yoon, and B. Birman, "Does Professional Development Change Teachers' Instruction? Results from a Three-Year Study," *Educational Evaluation and Policy Analysis*, **24**(2) (2002) 81–112.

[2] B.J. Fishman, R.W. Marx, S. Best, and R.T. Tal, "Linking Teacher and Student Learning to Improve Professional Development in Systemic Reform," *Teaching and Teacher Education*, **19** (2003) 643–658.

[3] T. Guskey and D. Sparks, "Linking Professional Development to Improvements in Student Learning," in E.M. Guyton and J.R. Dangel (eds.), *Research Linking Teacher Preparation and Student Performance: Teacher Education Yearbook XII*, Kendall/Hunt, Dubuque, IA, 2004.

[4] S. Loucks-Horsley and C. Matsumoto, "Research on Professional Development for Teachers of Mathematics and Science: The State of the Scene," *School Science and Mathematics*, **99**(5) (1995) 258–271.

[5] S. McDonald, J. Andal, K. Brown, and B. Schneider, *Getting the Evidence for Evidence-Based Initiatives: How the Midwest States Use Data Systems to Improve Education Processes and Outcomes*, Issues & Answers Report (REL 2007–No. 016), US Department of Education, Regional Educational Laboratory Midwest, Washington, DC, 2007; Internet: http://ies.ed.gov/ncee/edlabs.

[6] B.C. Clewell, P.B. Campbell, and L. Perlman, *Review of Evaluation Studies of Mathematics and Science Curricula and Professional Development Models*, GE Foundation, Urban Institute, Washington, DC, 2004.

[7] L.E. Suter and J. Frechtling, *Guiding Principles for Mathematics and Science Education Research Methods: Report of a Workshop, November 19-20, 1998*, National Science Foundation, Arlington, VA, 2000.

[8] P.H. Rossi, M.W. Lipsey, and H.E. Freeman, *Evaluation: A Systematic Approach*, Sage Publications, Thousand Oaks, CA, 2004.

[9] J. Baron, "Bringing Evidence-Driven Progress To Education MSP," Paper presented at the Regional Mathematics and Science Partnership Conference, The Council for Excellence in Government, The Coalition for Evidence-Based Policy, Washington, DC, 2006.

[10] *General Chemistry Brief Test for the Full-Year Course*, American Chemical Society, Division of Chemical Education, Examinations Institute, University of Wisconsin, Milwaukee, WI, 2006.

[11] *High School Chemistry*, American Chemical Society, Division of Chemical Education, Examinations Institute, University of Wisconsin, Milwaukee, WI, 2007.

[12] K.J. Roth, "Science Teachers as Researchers," in S.K. Abell and N.G. Lederman (eds.), *Research on Science Education*, Lawrence Erlbaum Associates, Mahwah, NJ, 2007.