DOCUMENT RESUME

ED 441 038 TM 030 837

AUTHOR	Lai, Morris K.; Young, Donald B.
TITLE	Comprehensive Standards-Based Data Collection: Essential for Valid Assessment of Program Impact.
PUB DATE	2000-04-27
NOTE	30p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Standards; *Data Collection; Elementary Secondary Education; *Evaluation Methods; *Inservice Teacher Education; *Professional Development; Program Evaluation; *Science Teachers
IDENTIFIERS	Impact Evaluation; University of Hawaii Manoa

ABSTRACT

To get a valid picture of program impact via comprehensive data collection, combined with a focus on standards, an evaluation of a professional development program for elementary and secondary level teachers of science was carried out at the University of Hawaii at Manoa. This paper presents selected examples from the program evaluation, designed to be both exceptionally comprehensive and focused on standards, and it discusses what the evaluators learned from the overall experience. The Standards-based Teacher Education through Partnerships (STEP) program was designed to empower teachers to become leaders in the standards-based movement. The program evaluation standards of the Joint Committee on Standards for Educational Evaluation were used to guide the evaluation design. Evaluators took a multidimensional assessment approach that included in-class case studies, videotapes of "best lessons," self-reports, in-class observations, student and teacher artifacts, teacher awards and recognitions, portfolios, performance tests, teacher-institute data, and student achievement data. Taking such a broad look at the program ensures that the essence of program impact is identified. Using multiple indicators of the achievement of project objectives allowed the triangulation and inclusion of what otherwise might be "fringe" indicators. Insights from the STEP program are summarized. (Contains 14 figures and 15 references.) (SLD)

Comprehensive Standards-Based Data Collection: Essential for Valid Assessment of Program Impact

Morris K. Lai & Donald B. Young

Curriculum Research & Development Group
University of Hawai'i at Mānoa

April 27, 2000 1

Paper presented at the Annual Meeting of the American Educational Research Association
Session 31.19: Distinguished Papers From State and Regional Research Association

Comprehensive Standards-Based Data Collection: Essential for Valid Assessment of Program Impact

When asked what skills the ideal evaluator should have, evaluation guru Michael Scriven replied that the answer was obvious—the ideal evaluator would possess all skills known to humanity. His response was not that facetious, for the quality of an evaluation could be enhanced by skills in diverse areas such as psychology, marketing, speech, forensics, computers, cognition, pedagogy, law, statistics, politics, and anthropology.

Similarly one could ask, "On what data would an ideal program evaluation be based?" A non-facetious response might be "An ideal program evaluation would include all relevant data." In today's educational environment, the ideal program evaluation at the K–12 level must also take into account program-evaluation standards and relevant subject-area standards.

In the spirit of getting to a valid picture of program impact via comprehensive data collection, combined with a focus on standards, we designed and carried out over a period of three years an evaluation of a professional-development program for elementary- and secondary-level teachers of science (Young, 1999). It was only after many, often onerous, data-collection efforts that we realized we could get a truly valid view of program impact only because we had completed a comprehensive collection of data, usually standards based but sometimes from the fringes (see Figure 1). Some of the data, when viewed alone, may seem to have relatively low generalizability or reliability; however, when viewed as one of many contributors to painting a valid picture of program impact, many of them acquired a degree of essentiality.

In fact, when we reflected on how we learn about various phenomena in life, we realized that such learning is often like the "little bit of" message in the popular song *Mambo No. 5*. For example, when one wants to evaluate one's efforts (or "program") to become a good/better tennis player, all of the following bits of data can be useful and essential: how one hits the ball in practice, won-loss record in league play, how close wins or losses were, how one feels physically and mentally after wins or losses, oral feedback from playing partner and opponents, the degree to which using different racquets or different strings made a difference, how much one looks forward to playing each week, whether one was able to improve in areas of weakness, articles in books or tennis magazines, and discussions with tennis-knowledgeable persons. A standardized pre-post tennis test would be far from adequate even with a "comparison group," even if such a thing were possible.

Evaluation Data Sources

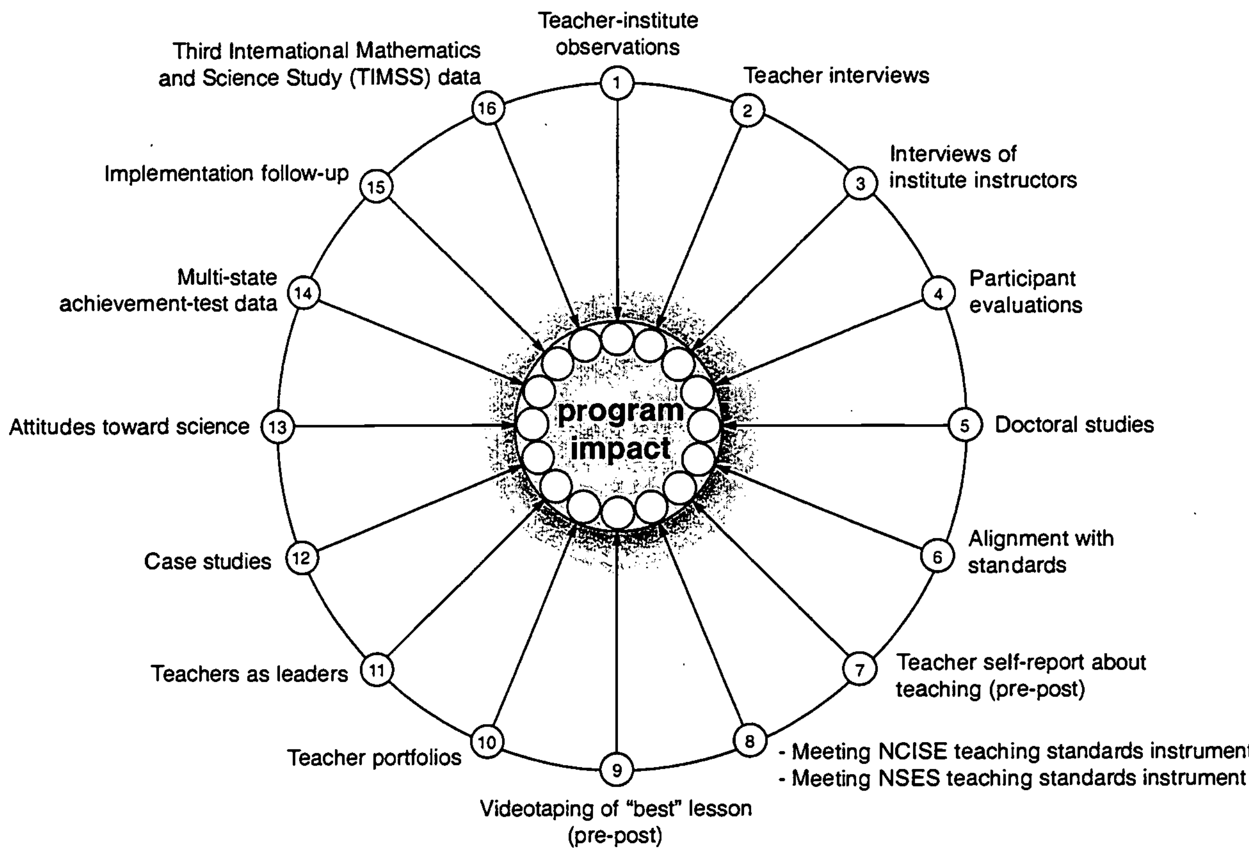

Figure 1. Many data sources can make essential contributions to the valid assessment of program impact (data sources 1–4, 6–9, and 12–15 were tapped through the use of standards-based data collection instruments).

In this paper we present selected examples from our program evaluation, designed to be exceptionally comprehensive and focused on standards, and discuss what we learned from the data collected and from the overall experience. We argue that other more commonly used, less comprehensive and less standards-based approaches are notably less able to validly assess program impact. Finally, we offer recommendations for how future evaluations might be best conducted, using such an approach.

The Project and the Players

The Curriculum Research & Development Group (CRDG) of the University of Hawai'i, in collaboration with 13 university partners and associated schools, created the Standards-based Teacher Education through Partnerships (STEP) project (funded by the U.S. Department of Education) to empower teachers to become leaders in the standards-based movement. STEP offered professional-development activities nationwide using the interdisciplinary science programs *Developmental Approaches in Science, Health and Technology* (DASH), *Foundational Approaches in Science Teaching* (FAST), *and Hawai'i Marine Science Studies* (HMSS). All three programs have been identified in the U.S. Department of Education's nationwide search as meeting the national standards for science education and professional development.

The timing of the STEP program evaluation presented a unique opportunity inasmuch as new standards for science education, professional development, and program evaluation had just been published. STEP was designed as a multi-year program, in which not only were multiple sites available, but also the sites were in several states providing opportunities to collect data from populations throughout the nation. We had the evaluation expertise of external independent contractors as well as university-based personnel at 13 different sites. Furthermore, we had the opportunity to develop and experiment with multiple indicators of impact including some that were standards based and others that were beyond the borders of most program evaluations.

Standards-Based Project Evaluation Design

We used the Program Evaluation Standards (Joint Committee on Standards for Educational Evaluation, 1994) to guide the comprehensive evaluation design. For the evaluation of the STEP Project, the 18 program evaluation standards (Joint Committee on Standards for Educational Evaluation, 1994) that were most directly relevant to this project evaluation were delineated.

Only recently have the final versions of nationally developed standards for science education been published (e.g., the *National Science Education Standards [NSES]* released by the National Research Council in late 1995 and published in 1996) for use by practitioners. Some associations have developed standards for the training of science teachers (e.g., the Association for the Education of Teachers of Science), some have posited principles or models (e.g., The National Institute for Science Education), some have developed frameworks based on the *NSES* (e.g., National Science Teachers Association's *A Framework for High School Science Education*

[1996]), and some have published benchmarks (e.g., American Association for the Advancement of Science, 1993).

Effect of Standards on Evaluations of Science-Education Programs

Not as obvious as the effects of science-education standards on teaching and curriculum development are the standards' effects on evaluation in the field. It is noteworthy to realize that if one buys into the science-education standards, then it directly follows that instruments developed for evaluating a specific program that are true to the standards should be essentially applicable to the evaluation of all other science-education programs. Just as is the case for curriculum development, however, certainly different versions of standards-based data-collection instruments can emerge.

To guide the overall evaluation we also used the U.S. Department of Education's Program Effectiveness Panel's submission guidelines manual (Ralph & Dwyer, 1988), *Emerging Roles of Evaluation in Science Education Reform* (O'Sullivan, 1995), and several recent documents on standards for science education and professional development.

To provide objective oversight of the evaluation, Jane Butler Kahle, Conduit Professor of Science Education at Miami University of Ohio, and Richard Shavelson, Dean of Stanford's School of Education, were contracted to review the design, instruments, methodology, and data-analysis procedures. We used their critiques to revise and fine-tune the evaluation.

Here we show a listing of the major relevant standards found in the recent literature. The evaluation data sources (ES__) in the second column are described subsequently.

TABLE 1. *Data Sources Addressing Various Standards*

Source of Standards	**Evaluation Data Sources**
Education Department general administrative regulations (*EDGAR*), 1994 revision	Overall STEP evaluation design for compliance
Program Evaluation Standards (AERA, NEA, & many others), 1994	Overall STEP evaluation design
American Association for the Advancement of Science (AAAS), *Benchmarks for Science Literacy* (11/93) [also *Science for All Americans*, 1989]	ES 4, ES 6, ES 10, ES 13, ES 16
National Center for Improving Science Education (NCISE) [in *Promising Practices*]	ES 2, ES 3, ES 4, ES 8, ES 10, ES 15
National Research Council, *National Science Education Standards*, 1996. Teaching, Professional Development, and Assessment.	ES 1, ES 2, ES 3, ES 4, ES 5, ES 6, ES 7, ES 8, ES 9, ES 10, ES 11, ES 13, ES 14, ES 15, ES 16
National Staff Development Council (NSDC), *Standards for Staff Development*, 1994	ES 1, ES 2, ES 3, ES 4, ES 6, ES 7
Program Effectiveness Panel (U.S. Dept. of Education) guidelines (e.g., as outlined in *Making the Case*, 1988)	ES 5, ES 10, ES 11, E 12, ES 13, ES 14, ES 15, ES 16
Other Standards (less emphasis on)	E.g., U.S. Dept. of Education (1994 draft) and ASCD (11/94 draft).

The ESs refer to the following delineated list of 16 evaluation data sources that constituted the essential elements of the evaluation.

STEP Evaluation Data Sources

ES 1. Teacher-Institute Observations

- External evaluator's observations of sample of teacher institutes using instrument based on the professional development standards and rating scale
- Identify random sample of institutes
- Make observations on random sample of institute days

ES 2. Teacher Interviews

- External evaluator interviews of teacher participants to verify observations from ES 1
- Interviews to assess degree of implementation of standards not observed

ES 3. Institute Instructor Interviews

- External interview of institute instructors to verify observations from ES 1
- Interview to assess degree of implementation of standards not addressed
- Interviews to compare lesson objectives with observations

ES 4. Participant Evaluations

- Collect participant evaluations on standards-based instruments at the end of institutes
- Two instruments, one Likert-type response, one open response
- Review and summarize participant evaluation data from summer by institute and by state

ES 5. Doctoral Studies

- Assist doctoral researchers in accessing data
- Focus on degree of implementation
- Teacher observations and interviews

ES 6. Alignment with Standards

- Analyze and document alignment of program content with AAAS Benchmarks
- Analyze and document alignment of program content with NSES
- Analyze and document alignment of professional development strategies with NSES and NSDC standards

ES 7. Teacher Self-Report about Teaching

- Develop instrument based on NSES and/or NCISE to measure impact on teaching
- Administer before project involvement
- Administer at the end of subsequent academic year

ES 8. Meeting NCISE and NSES Teaching Standards

- Develop standards-based free-response instrument
- Administer at the end of each institute

ES 9. Videotape "Best" Lessons

- Select teachers in elementary and middle school
- Protocols for consistent videotaping including selection of classes
- Collect demographic data
- Videotape "best" class before project involvement and again after one academic year
- Analyze for evidence of achieving NSES teaching standards

ES 10. Teacher Portfolios

- SES protocols for developing portfolios
- Identify elementary and middle school teachers who are high implementers
- Analyze for evidence of achieving content and teaching standards

ES 11. Teachers as Leaders

- Collect and catalog indicators of developing leadership. (Awards, anecdotes; professional meetings attended; presentations at professional meetings; within school faculty development; becoming certified trainers; action research reports; supervisor testimonials; enrolling for advanced degree; school, district, state, or professional committees; publications; newspaper reports; other)
- Survey institute participants
- Compile data and report

ES 12. Case studies

- Conduct case studies in elementary and middle school implementations
- Contract external evaluator to conduct cross case analysis
- Analyze for evidence of achieving NSES standards

ES 13. Attitudes Toward Science

- Develop student survey instrument
- Administer to elementary students
- Correlate responses with degree of implementation of standards-based program

ES 14. Multi-State Achievement Test Data

- Collect existing student impact data with comparison groups (standardized test data by class/school; school demographics; comparison data; performance testing; reading/mathematics scores; other indicators of impact)
- Categorize, analyze, and report

ES 15. Implementation Follow-up

- Collect data from multiple sources on success of implementation including classroom observations, teacher interviews, administrator interviews, teacher feedback on survey instruments, teacher meetings, and other sources.
- Triangulate data to verify validity.

ES 16. Third International Mathematics and Science Study

- Collect and analyze student impact data available through international comparisons
- Analyze students' achievement in standards-based classes with comparison groups

Inappropriateness of some standards for developing items for immediate feedback

In our attempt to directly address some of the standards, we found that some were longitudinal in nature or were to be addressed at the teachers' schools rather than at a staff-development institute. Such standards stimulated us to augment the evaluation in a couple of major ways. We had teachers videotape their "best lesson" before the institute. We asked them to do the same a year later. An external expert using science-education teaching standards from the National Research Council then analyzed these videotapes. We also had teachers fill out before the start of the institute a Likert-type scale addressing how they taught in their classroom. We then gave them the same questionnaire during the end of the subsequent school year.

A Sampling of Data-Collection Efforts

To illustrate the method being advocated, we will focus on the project's institutes and related activities, teachers' use of standards-based pedagogy and content, and students' demonstration of learning. Table 2 presents a matrix showing the numerous data sources that were used to investigate those areas.

TABLE 2. *Evaluation Data Sources for Major Project Areas*

Project Area	Program Evaluation Data Sources															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Institutes and activities meeting or exceeding standards	√	√	√	√												
Participants using standards-based pedagogy and content				√	√	√	√	√	√	√	√	√	√			
Students (taught by STEP participants) demonstrating mastery of concepts												√		√	√	√

Data source key:

1. Teacher-institute observations
2. Teacher interviews
3. Interviews of institute instructors
4. Participant evaluations
5. Doctoral studies
6. Alignment with standards
7. Teacher self-report about teaching (pre-post)
8. Meeting National Center for Improving Science Education (NCISE) or National Research Council (NRC) teaching standards instruments
9. Videotaping of "best" lesson (pre-post)
10. Teacher portfolios
11. Teachers-as-leaders indicators
12. Case studies
13. Attitudes-toward-science questionnaire
14. Multi-state achievement-test data
15. Implementation follow-up
16. Third International Mathematics and Science Study (TIMSS) data

Data-collection instruments generally flowed from the standards but often required a careful reading and interpretation of the standards. With some modest modifications of wording and scope, we were able to produce standards-based instruments such as Likert-type self-report rating scales, open-ended interview schedules, classroom and teacher-training observation instruments, and overall program-evaluation checklists. We now present specific examples of this approach incorporating standards into data-collection instruments.

Institutes and Activities

Four data sources were used to help assess the degree to which institutes and activities met standards: (a) direct observation, (b) interviews, (c) participant feedback, and (d) doctoral study.

Direct Observations and Interviews.

Observational Data Collection During Training Institute

We developed an observation instrument based on professional-development standards. As expected, it was somewhat more difficult to develop an instrument in this area. The instrument we developed for the teacher institutes had observers rate standards (e.g., as follows: "Observed—Clear focus; Observed—Adequately addressed; Observed—Somewhat addressed; Not observed").

The *Observations of STEP Institutes Summer 1995* was used in three ways. First, the external evaluators observed random training sessions—three half-day sessions in three different DASH institutes; twelve half-day sessions in two different FAST institutes; four half-day sessions in two different HMSS institutes. Second, the external evaluators interviewed randomly selected participants asking questions about those standards they were unable to observe. Third, the external evaluators interviewed the instructors, again with the emphasis on those standards that had not been observed. The activities were conducted in that order by each external evaluator to avoid contamination of the original observations.

The three external evaluators used the instrument in crossover tests to establish acceptable inter-rater reliability levels before observations began. Evaluators, instructors, and participants gave their opinion of the degree to which each standard had been addressed by choosing one of the following: (a) clearly addressed, (b) adequately addressed, (c) somewhat addressed, (d) not addressed, or (e) not observed/experienced.

The National Science Education Standards include professional-development standards for science education, which we used to create an observation instrument that had 21 items. The evaluators used four categories to rank the observations. Working definitions of each category are shown in Table 3.

The external evaluators observed training sessions as follows: one full day of one K–3 DASH institute; one full day of one grade 4–5 DASH institute; six half-day sessions of 3 FAST institutes. The HMSS institutes were not observed in Year 2 due to a scheduling error on the part of the external evaluators.

TABLE 3. *Rankings and Definitions for the Observation of STEP Institutes Instrument 1996*

Ranking Category	Working Definition
Observed: Clear Focus	This ranking denotes that the evaluator observed the element. The instructor clearly focused on the element, either by addressing it at several different points in the observed presentation or by an extended discussion or demonstration of the element at a single point. The essential content, point, or purpose of the element was thoroughly communicated to participants.
Observed: Adequately Addressed	This ranking denotes that the evaluator observed the element. The instructor addressed the element at some point in the observed presentation. The essential content, point, or purpose of the element was communicated to participants.
Observed: Somewhat Addressed	This ranking denotes that the evaluator observed an element. The instructor addressed the element at some point in the observed presentation. A part of the essential content, point, or purpose of the element was communicated to participants.
Not Observed	This ranking denotes that the evaluator did not observe the element.

Participant Feedback: CRDG science-teacher institute. Every CRDG science-teacher institute is assessed for quality using a 5-point Likert scale, which includes items addressing (a) specific curriculum content, (b) quality of the workshop, and (c) major staff-development standards. We adapted this instrument to include items from the most recent professional-development standards. The data showed that the institutes effectively met the professional-development standards to a high degree and were consistent across years and across programs (see Table 4 and Figure 2).

TABLE 4. *Summary DASH Institute Data on Addressing Professional Development Standards*

Questions	Mean 1995	Mean 1996	Mean 1997	Mean 1998
1. The institute included theory, demonstration, practice and coaching.	4.8	4.8	4.8	4.8
2. The institute was conducted in a learning climate that was collaborative, informal, and respectful.	4.8	4.8	4.9	4.9
3. The institute increased my ability to provide a challenging, developmentally appropriate curriculum based on desired skill and knowledge outcomes for all students.	4.6	4.6	4.8	4.7
4. The institute prepared me to demonstrate high expectations for student learning.	4.5	4.5	4.6	4.5
5. The institute improved my ability to engage parents and families in improving their children's educational performance.	4.1	4.1	4.2	4.1
6. The institute prepared me to use an evaluation process that is ongoing, includes multiple sources of information, and focuses on all learners.	4.4	4.3	4.5	4.4
7. The institute increased my understanding of how to provide school environments and instruction that are responsive to the developmental needs of students.	4.7	4.5	4.6	4.6
8. The institute enhanced my ability to have students exercise the meaningful application of knowledge.	4.4	4.6	4.7	4.7
9. The institute prepared me to use research-based teaching strategies appropriate to my instructional objectives and my students.	4.5	4.4	4.6	4.5
10. The institute enhanced my ability to provide an equitable and quality education to all students.	4.6	4.5	4.6	4.5
11. The institute helped me learn and apply collaborative skills to work collegially with others.	4.3	4.6	4.7	4.6
12. The institute prepared me to develop and implement classroom-based management plans that maximize student learning.	4.3	4.4	4.5	4.4
Number of respondents	895	1,010	687	380
Number of institutes sampled	68	90	55	42

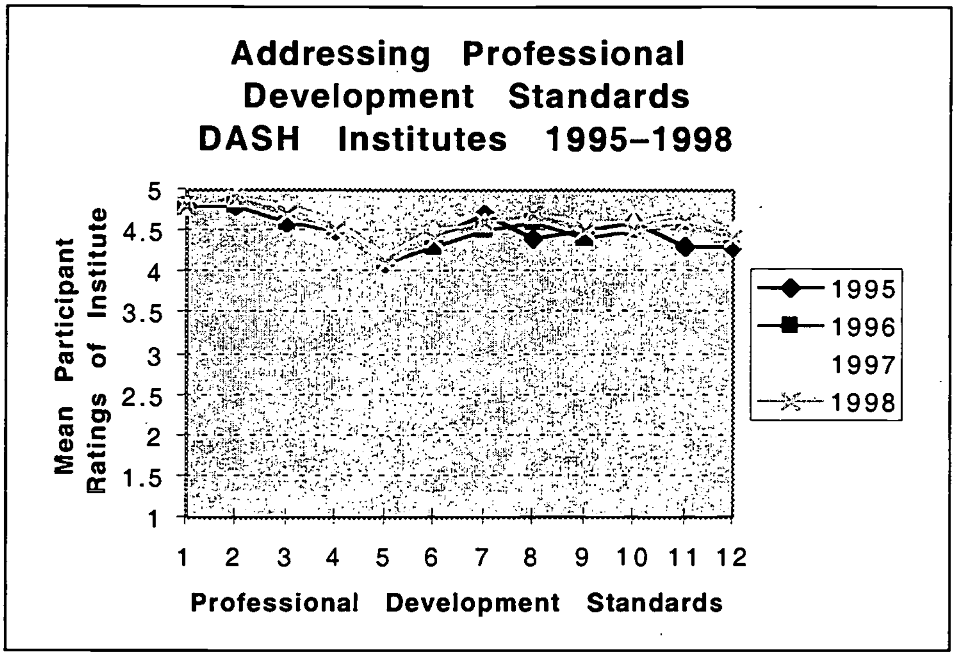

Figure 2. Teacher ratings on professional-development standards for DASH institutes 1995–1998.

Doctoral Study. In a doctoral study conducted in 1992–1993 at the University of Kansas, Kesner (1993) investigated the impact of the STEP professional-development activities in elementary schools in a suburban Kansas school district. At the end of nine months, the seven teachers studied were all at the routine level of use or above (Fuller, 1969; Hall, 1979; Hall & Loucks, 1978; Hord & Huling-Austin, 1987), indicating that teachers moved rapidly through the stages of concern and that the professional-development activities had positive effects on implementation (see Figures 3 and 4).

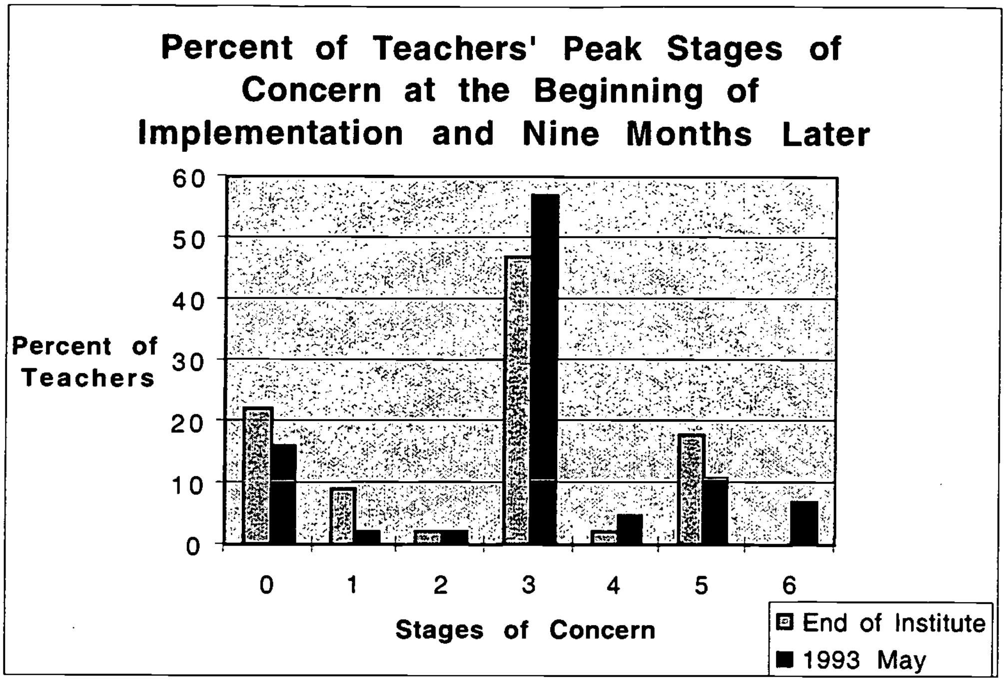

Figure 3. Shift in teacher stage of concern with DASH professional development ($n = 43$).

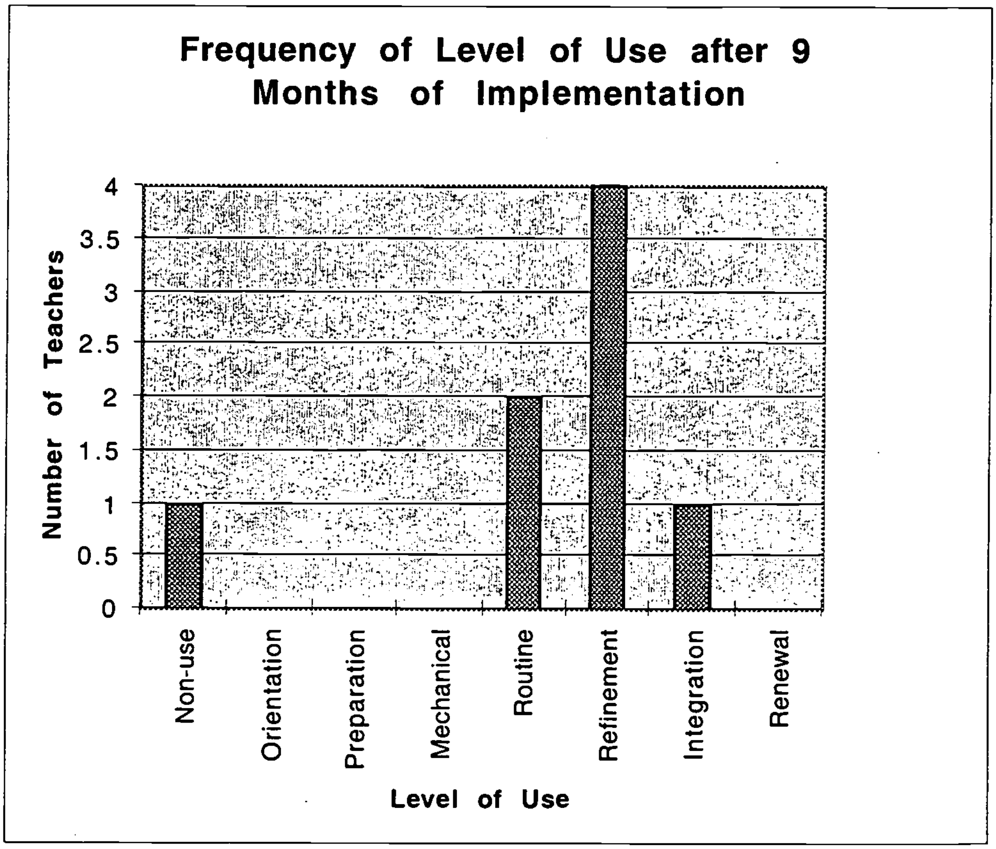

Figure 4. Success of implementation after 9 months with professional development and support ($n = 8$).

The data collected on the four indicators provided information on how well the institutes addressed the standards and how well the instructors taught at the institutes.

Participants Using Standards-based Pedagogy and Content

Indicators selected to evaluate participants' use of standards-based pedagogy and content included (a) alignments of programs with teaching standards, (b) teacher self-report questionnaires, (c) videotapes of teachers' "best" lessons, (d) teacher portfolios, (e) teachers-as-leaders data, and (f) case studies.

Taken collectively, the content alignment of the STEP programs with the NSES, American Association for the Advancement of Services (AAAS) Benchmarks, and selected state science frameworks; the evidences that teachers know and can give examples of the science teaching standards on an open-ended instrument; the teacher self-reports on changes in their teaching; the videotape documentation of classroom teaching; the teacher portfolios; the list of leadership activities in which teachers have engaged; the case studies; and the Carnegie Mellon institute and follow-up data provide strong evidence that participants used standards-based pedagogy and content most of the time.

Best Lessons. Selected teachers were asked to videotape their "best science lesson." An independent, external evaluator, using the Instrument for the Observation of Teaching Activities (IOTA) [National IOTA Council, 1970], analyzed the videotapes. In an attempt to economize, we asked teachers to set up video equipment in their classrooms according to protocols provided. In retrospect, this created significant limitations on the utility of data. The cameras were stationary, focused on one section of the classroom and students. Verbal interactions were difficult and, in some cases, impossible to hear or interpret.

Especially noteworthy was the consistently higher performance of teachers who had experienced STEP professional-development activities in IOTA categories such as Variety in Learning Activities, Use of Materials of Instruction, and Opportunity for Participation. Teachers participating in STEP professional-development activities generally did less well on Learning/Interest Centers and Individualized Instruction, areas not emphasized in program strategies (see Figures 5 and 6).

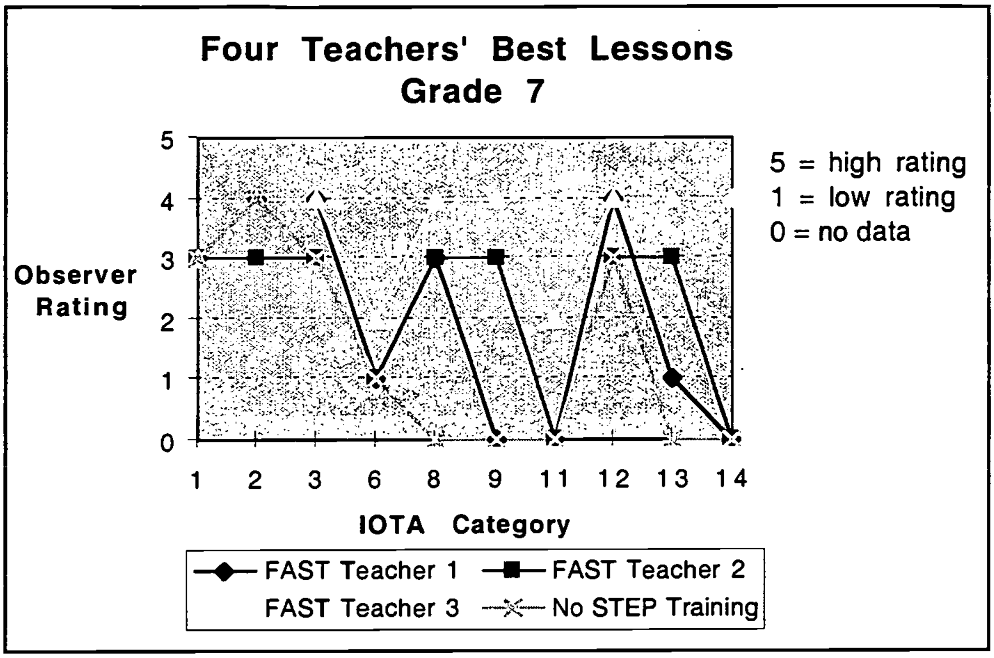

Figure 5. IOTA scaling of "best lesson" prior to the FAST institute and 11 months later after participating in FAST professional-development activities (four different teachers; three participated in FAST).

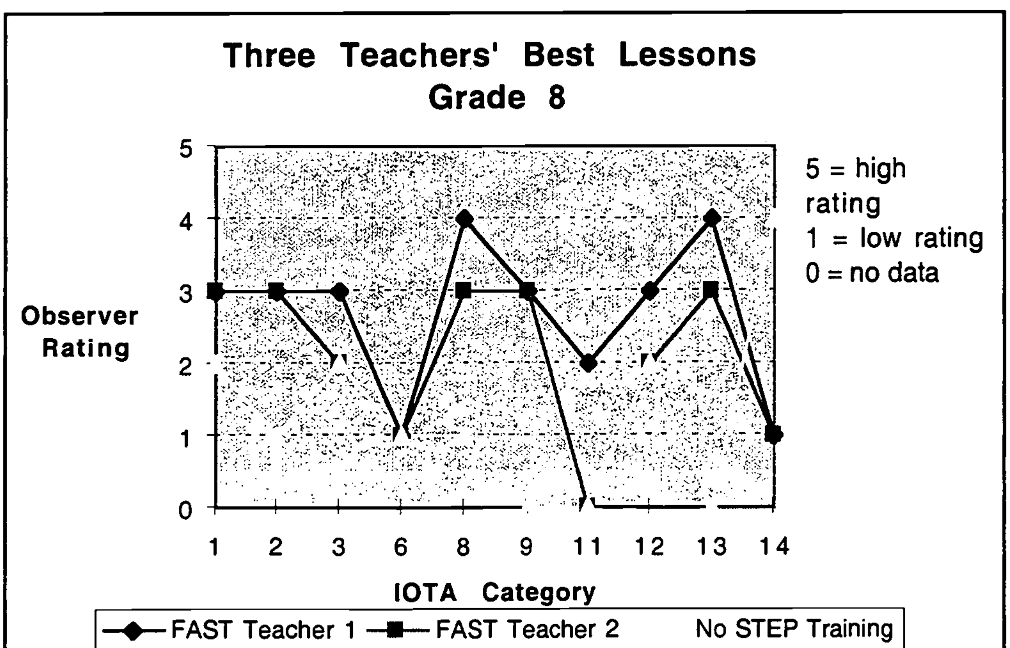

Figure 6. IOTA scaling of "best lesson" prior to the FAST institute and 11 months later after participating in FAST professional-development activities (three different teachers; two participated in FAST).

We believe that the idea of obtaining videotapes of teachers' "best lessons" is a good one, but we need to (1) provide better protocols that stipulate that teachers should clearly establish the lesson objectives so that the evaluator can identify them, (2) use multiple cameras and microphones to get better quality data, and (3) assign staff or hire others to do the videotaping.

Teacher Portfolios. Twelve selected teachers were asked to prepare portfolios addressing specific issues. For the purpose of this paper, we present a couple of sample comments from the portfolios.

> *I have clearly experienced the quality education that can emerge when teachers are able to plan and implement learning that builds upon prior knowledge and experiences deliberately sequenced across grade levels.* This teacher has 15 years of experience and has provided strong leadership in two elementary schools.
>
> *My experiences working with teachers from elementary schools throughout the state of Hawai'i have led me to believe very strongly in the need for curriculums like DASH.*

Sample parent comments taken from the portfolio regarding the use of DASH in a school-within-a-school setting:

> *Not only have we seen a great improvement in the development of higher order thinking skills, but also present are: the student sense of belonging, student attitudes toward school in general and particular subjects, social bonding between teachers and students, hands on learning, and cooperative learning. More importantly, the personal growth of my son has been remarkable....Once a quiet and shy student, he has become very confident and outspoken, unafraid to communicate his ideas either verbally or in writing.*

Teachers as Leaders. One of the goals of STEP was to develop teacher leaders in the science-education reform movement. We took a broad definition of leadership to include things such as awards, attendance and presentations at professional meetings, becoming certified trainers, conducting action research, seeking or completing an advanced degree, and publications. A sampling of findings for project participants follows.

National Award Winners

12 Presidential Award Winners for Excellence in Science Teaching

7 National Association Award Winners

11 State Award Winners

3 National Board Certified Teachers

Doctoral programs

12 STEP participants

Masters programs

11 STEP participants

Additional Indicators of Leadership

- Approximately 95 project teachers participated in science-education conferences for the first time
- Approximately 53 project teachers presented at science-education conferences for the first time

Case Studies (focus on teachers). Fourteen in-depth case studies were conducted in DASH and comparison classrooms. Results show that experienced teachers using DASH increased emphasis on science and improved their focus on student learning. DASH teachers spent more time teaching science than previously and used a richer set of strategies. They integrated science with other subjects more effectively than they did in the past. These teachers consistently report that DASH has given them renewed enthusiasm for teaching science.

Two doctoral candidates at Stanford University (Gilroy, 1995; Shih, 1998) conducted independent case studies of the impact of FAST in San Jose, CA schools. These studies found FAST to have a major impact on teachers' instructional strategies, the curriculum, and expectations of students. Some excerpts follow:

> *How well prepared do students feel upon entering the traditional classes?*
> The students generally feel very confident when going into biology from the FAST class. Their experience with discovery has given them confidence that they can appreciate the depth in a given field.
>
> *How do colleges view the FAST curriculum?*
> The local university, San Jose State, has been an active partner in sculpting the implementation and adaptation of FAST, and is quite happy with the program.

Do teachers feel FAST is complete?
The teachers all spoke highly of the program. They feel it recognizes realities of the learning process that have traditionally been ignored.

Students Demonstrating Mastery of Concepts

All STEP activities were directed toward teachers and improving instruction in science; however, these efforts are ultimately intended to positively affect student achievement. Although the STEP evaluation team did not attempt to test student performance directly, we did collect student-impact data through case studies and available achievement-test data.

Case Studies with Focus on Students. Fourteen case studies showed that, during DASH activities, students consistently demonstrated a high proportion of engaged learning time (time on task). The data also showed that students connected and applied what they learned in school to their lives outside of school. They also showed increased curiosity and proficiency in inquiry skills including questioning, observation, measurement, and use of instruments and information sources to acquire data.

Studies by Doctoral Students. Gilroy (1995) studied the impact of FAST on student learning. Some excerpts:

Has FAST increased student interest in science?
Before FAST... the school offered 1 physics and 3 chemistry classes. Now demand has increased that to 3 physics and 10 chemistry classes. Students explicitly identified the format of the 9th grade class (FAST) as making science approachable and interesting.

Have teachers noticed a significant difference in performance?
There are no students who begin . . . without FAST, but there are a significant number who transfer later. Despite coming from programs more in tune with traditional format of the upper level classes, these students seem (in general) less able to fit in. With effort to give them the skills that FAST inculcates, they often find their feet and adapt.

How many students go on to college? Has this increased due to FAST?
Approximately 35% of the senior class . . . go on to attend a college. A number of these apply to and attend the most challenging technical and science-oriented schools. The anecdotal evidence is that this number has climbed sharply since the introduction of FAST.

In a 1998 study in the same district, Shih (1998) used student interviews to examine the impact of the FAST experience on student learning.

How was the FAST program different or similar compared to science classes you took in the past?
Unlike other classes, we got to actually get to use stuff in it. They [teachers] thought we were responsible enough to use the items.

What values, abilities, or attitudes did you gain or build upon from FAST?
I learned to have an open mind, to think more, take different things into perspective. Not just one way. I think it was the labs. Because with the hypothesis you think, what's going to happen and stuff, and then it turns out a different way.

Attitudes Toward Science. A study of the impact of DASH on attitudes toward science was conducted by researchers at the University of Missouri at St. Louis in 1995–96. The conclusion from this study was that the more a teacher uses DASH, the more positive were students' attitudes toward science (see Figure 7).

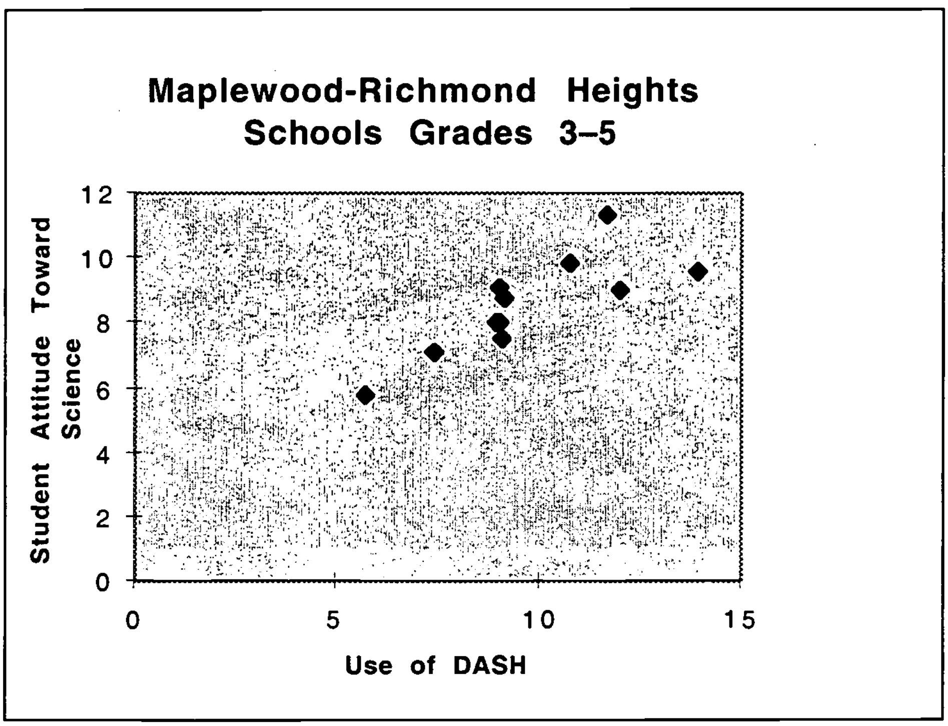

Figure 7. Maplewood-Richmond Heights School student attitude toward science.

Achievement-Test Data. Standardized achievement-test data were collected from schools in eight states. As an example, the 1998 science-achievement results for Missouri's

Mastery and Achievement Tests, as reported for St. Louis public schools, showed that students using DASH achieved above the state norms in all sections. School officials credit DASH with these student-achievement results.

Westport, CT Achievement-Test Data: Westport School District in Connecticut has used FAST as its middle school science program since 1990. Recently, with the movement toward standards-based education and accountability, Connecticut introduced its own version of content and performance standards for all subject areas and followed up with the development of the Connecticut Academic Performance Test (CAPT) to measure student achievement in science, mathematics, interdisciplinary, and language arts. The CAPT was administered for the first time in 1995 to all tenth-grade students. Data in 1996 and 1998 show sharp increases in the CAPT science scores, which have been maintained well above state, expected scores. The data reflect achievement of students who have had FAST as their middle-school science experience and indicate that the science program is preparing students well for such standards-based measures of achievement. Figure 8 graphically shows these data. Scores for 1997 were not available to us.

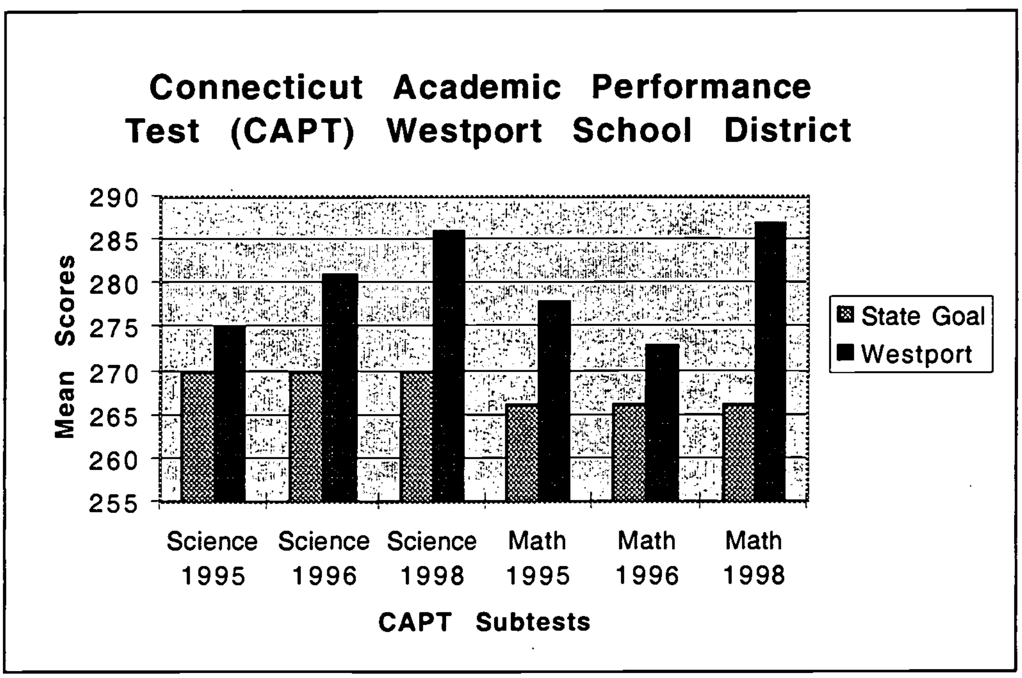

Figure 8. Westport school district performance on CAPT 1995-1998 for grade 10.

Mililani, HI Achievement-Test Data: At Mililani Uka Elementary School data were collected on the Environment subtest of the Stanford Achievement Test (SAT). The SAT reading and mathematics battery are administered to all schools on a yearly basis in Hawai'i. Mililani Uka is a large (1,200 students) suburban school on O'ahu serving a middle-class community of diverse ethnic mix. DASH is the main science curriculum of the school. Figure 9 shows the stanine distribution of one class of grade 2 students who had experienced 2 years of DASH on the SAT reading, mathematics, and environment subtests. The teacher attributes the students' mathematics and science performance to the DASH experience. Reading scores, on the other hand, were below national norms.

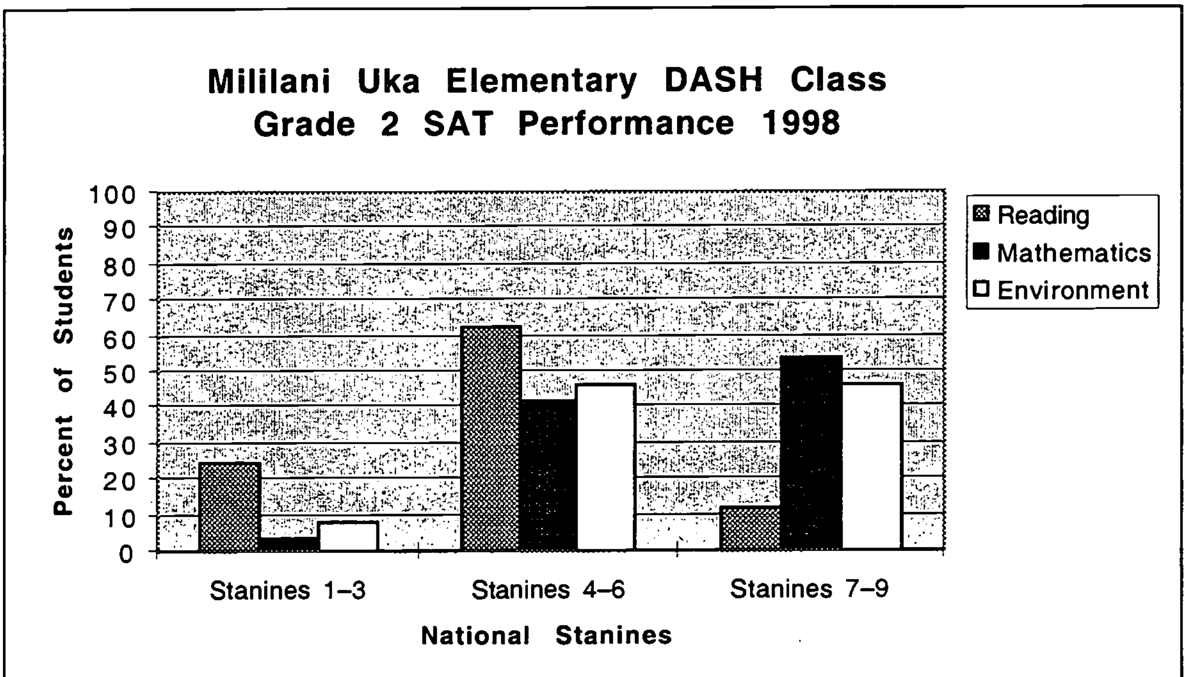

Figure 9. DASH grade 2 student achievement on SAT subtests at Mililani Uka Elementary School.

Kailua, HI Achievement-Test Data: Ka'elepulu School on O'ahu became one of the first school-community-based management schools in the state in 1989. As its core curriculum the faculty selected DASH and participated in the professional-development institutes and follow-up activities. Science scores on the Stanford Achievement for grade 3 students at Ka'elepulu were noticeably higher than those of the district or the state.

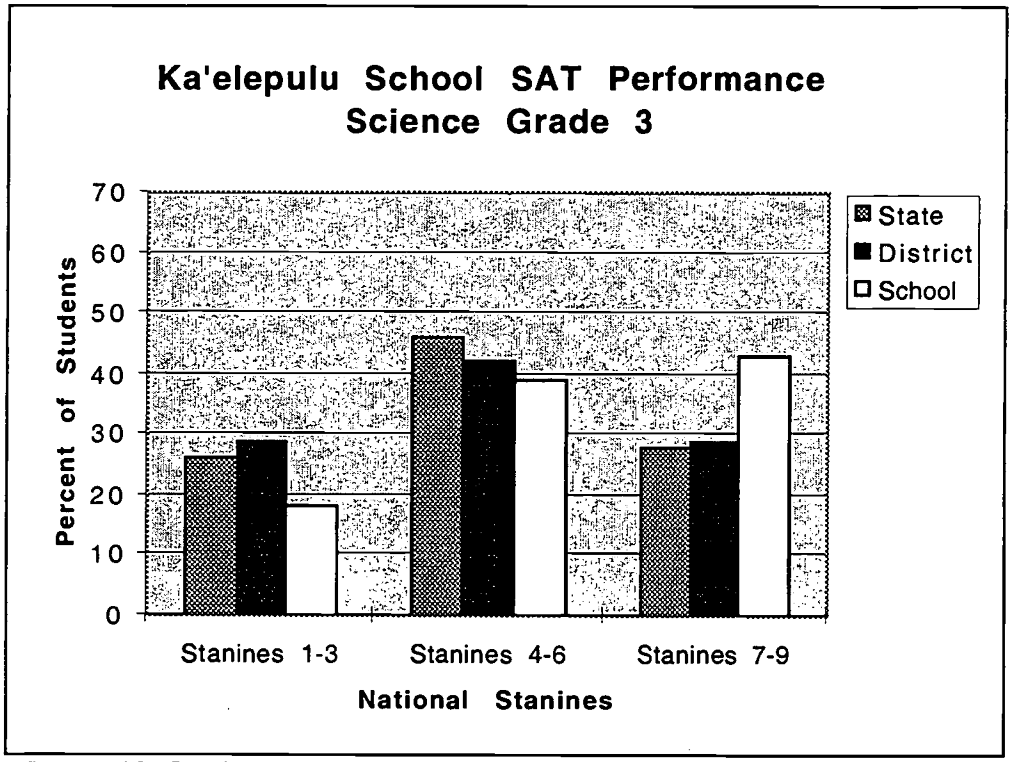

Figure 10. Grade 3 student performance on SAT science subtest.

International Study (TIMSS). For the past six years, CRDG staff have worked with science educators in Slovakia to translate and introduce FAST. In 1996 students who had completed two years in FAST were tested as part of the Third International Mathematics and Science Study (TIMSS). A comparison of the achievement of students learning science in FAST and a representative sample of the national population educated in the classical way was used as an assessment of impact. FAST students consistently scored significantly higher than the Slovakia national average (see Figures 8–11). Local (Slovakian) educators informed the STEP Project that the Slovakia FAST students were demographically similar to the national Slovakia sample.

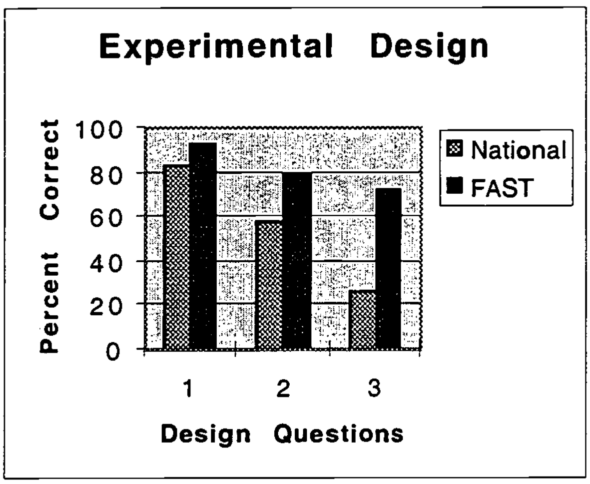

Figure 11. Slovakia FAST student performance vs. Slovakia national performance on experimental-design questions.

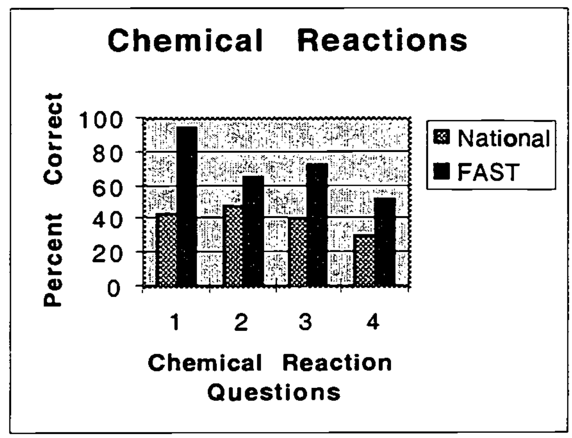

Figure 12. Slovakia FAST student performance vs. Slovakia national performance on chemical-reaction questions.

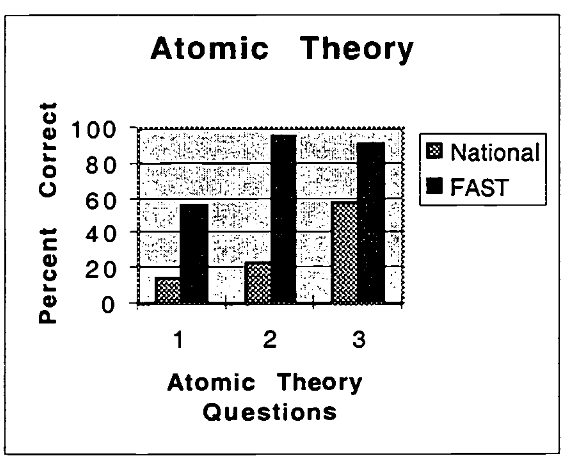

Figure 13. Slovakia FAST student performance vs. Slovakia national performance on atomic-theory questions.

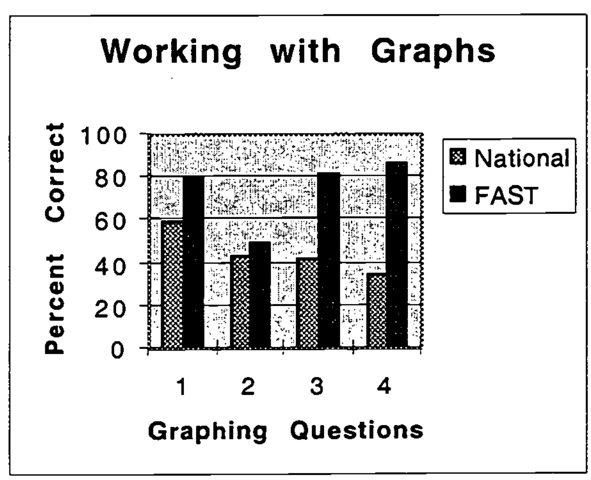

Figure 14. Slovakia FAST student performance vs. Slovakia national performance on graphing questions.

Discussion and Conclusions

The program evaluation used to determine the impact of STEP's efforts was designed to be comprehensive using multiple indicators, some of which can be considered beyond the boundaries of traditional approaches. We took a multidimensional assessment approach that included in-class case studies, videotapes of "best lessons," self-reports, in-class observations, student and teacher artifacts, teacher awards and recognitions, portfolios, performance tests, teacher-institute data, and student-achievement data. Whenever feasible we directly used the National Science Education Standards and National Staff Development Council Standards to design our data-collection instruments.

If we had not taken such a broad sweep, we could not be confident that we had identified the essence of the impact of the project. Any one of the indicators in and of itself might not provide convincing data on impact. However, taken collectively the multiple indicators paint a convincing picture of positive change that goes beyond the classroom. For example, having data on teachers-as-leaders provides insight into the project's effects on teachers' professional activities, some of which are outside of school.

From the broad landscape of indicators we know, for example, that teachers who participated in project activities know and understand the NSES standards; teachers can cite quality examples of how they can meet those standards; teachers changed instructional strategies in ways that are consistent with the standards; teachers emerged as leaders in the standards-based reform effort.

Regarding the professional-development efforts, we can point to the strategies that were effective in achieving desired changes in teacher behaviors. We know that the professional-development activities provided were consistent with current professional-development standards. We know that the strategies and activities used were consistent over time. We know how instructors used formative data to adjust their teaching in ways that better met the professional-development standards.

At the classroom level, we know from achievement-test data and other indicators that the professional-development activities that changed teacher behaviors had a positive impact on students' learning. The conclusion we draw is based on different tests administered to students in different classrooms in different states and as well as a different country. The overall picture is one of positive impact on learning.

Using a small number of indicators, even selecting the "best" (small) set of potential ones, would be insufficient, questionable, and not likely to enable us to draw the conclusions about project impact that we can now confidently do. For example, if we had used only external observers in a sampling of teacher institutes, we would not be able to clearly document alignment of the professional-development activities with the standards. Similarly, if we had relied mainly on case studies to determine the impact of project activities on teaching, we could not be so confident in generalizing from the data. If we ourselves had tested a sample of students directly, we could not be as confident as we now are about the impact on student learning. The consistent finding of positive impact across multiple sites using multiple indicators of student achievement is in our opinion more convincing. Also by obtaining findings from independent doctoral studies conducted under the guidance of faculty who had no vested interest in the STEP program, we had further strong corroboration of the more in-house findings.

By using multiple indicators regarding the achievement of project objectives, we were able to triangulate and include what might otherwise be "fringe" indicators. When the data from these indicators were combined, they reinforced one another and enabled us and project staff to gain new insights into the broader impact of project activities. In addition, we gained new insights into which indicators are most reliable, give the best evidence of impact, and are most cost efficient, thus enabling us to refine the design for future applications. The many "fringe" indicators add up to the point that their contribution is essential (see Figure 1). The accumulation of a little bit of lots of information reflects how learning often takes place in life.

Among the insights we now have are these:

Regarding indicators of teacher impact,

- Systematically collecting teachers-as-leaders data can demonstrate long-term impact.
- Institute observations and interviews were no more effective indicators than a combination of teacher standards-based, self-report instruments. (Such standards-based instruments should be appropriate to use in most evaluations of staff-development programs for science teachers).
- Videotaping "best" lessons is an effective indicator but will require an external film crew and rethinking the use of IOTA.
- Teacher portfolios may be a useful tool to evaluate impact but may be difficult to obtain.
- Given the barriers to implementation of standards-based reform, case studies, although expensive and time consuming can provide excellent data on actual classroom use.

Regarding indicators of student impact,

- Achievement-test data on different tests collected over multiple sites, can serve as an excellent indicator of impact on student learning.
- Studies done by doctoral students can provide valuable research information that was independently obtained and under the quality control of graduate-school faculty.

We advocate thinking broadly in designing an evaluation and including multiple indicators, some of which may be on the fringes to provide a more essential picture of impact. Only by first carrying out such comprehensive data collection can evaluators hope to get to the essence of the impact of any program, and thereby to the essence of the evaluation.

Although the approach we used may seem less systematic than is normally the case in program evaluation, this should not necessarily feel discordant because our ways of knowing and learning are akin to such an approach. We learn about many things in life through diverse, often unsystematic, activities such as reading, trying, observing, asking, discussing, sharing, and listening to little bits at a time. One could argue that each interaction produces some additional learning. Many (perhaps all) interactions may prove to be essential to reaching the current state of understanding. An evaluation that leads to the highest level of understanding of a program and its impact is in our minds the best kind of evaluation possible.

References

American Association for the Advancement of Science. (1989). *Science for all Americans*. Washington, DC: Author.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. Washington, DC: Author.

Fuller, F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6*, 207–226.

Gilroy, B. (1995). *Education 359A: Summary and evaluation of interfaces in FAST*. Unpublished paper, Stanford University, Stanford, CA.

Hall, G. (1979). The concerns-based approach for facilitating change. *Educational Horizons,* 57, 202–208.

Hall, G., & Loucks, S. (1978). Teacher concerns as a basis for facilitating and personalizing staff development. *Teachers College Record. 80*, 36–53.

Hord, S., & Huling-Austin, L. (1987). Effective curriculum implementation: Some promising new insights. *Elementary School Journal, 87*, 97–115.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.

Kesner, C. (1993). *A study of an elementary staff development process implementing the DASH science program.* Ann Arbor, MI: University Microfilms International.

National IOTA Council. (1970). *Instrument for the observation of teaching activities.* Tempe, AZ: Author.

National Research Council. (1996). *National science education standards.* Washington, DC: Author.

National Science Teacher Association. (1996). *A framework for high school science education.* Washington, DC: Author.

National Staff Development Council. (1994). *National Staff Development Council's standards for staff development: middle level edition.*

O'Sullivan, R. G. (Ed.). (1995). Emerging roles of evaluation in science education reform. *New Directions for Program Evaluation*, 65 (spring).

Ralph, J., & Dwyer, M. C. (1988). *Making the case: Evidence of program effectiveness in schools and classrooms*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Shih, W. (1998). *Objectives realized: A case study of the outcomes of the Foundational Approaches in Science Teaching (FAST) Project*. Unpublished paper, Stanford University, Stanford, CA.

U.S. Department of Education. (1997). *Education Department general administrative regulations.* Washington, DC: Author.

Young, D. B. (1999). *Standards-based Teacher Education Through Partnerships (STEP). Final performance report*. Grant Number R215J40238-96. Honolulu: University of Hawai'i, Curriculum Research & Development Group.

AERA

ERIC®

TM030837

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Comprehensive Standards-Based Data Collection: Essential for Valid Assessment of Program Impact	
Author(s): Morris K. Lai & Donald B. Young	
Corporate Source:	Publication Date: 4/27/00

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ____ Sample ____ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please →

Signature: Morris K. Lai	Printed Name/Position/Title: Morris K. Lai, Director of Evaluation	
Organization/Address: Curriculum Research & Development Group, University of Hawaii, 1776 University Ave., Honolulu, HI 96822-2463	Telephone: 808-9567900	FAX: 808-956-9510
	E-Mail Address: lai@hawaii.edu	Date: 4/27/00

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20772
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com