# REVIEW OF NOAA EDUCATION PROGRAM EVALUATION REPORTS for the Committee to Review the NOAA Education Program

**Prepared by Ann Brackett**
**Independent Consultant**
**April 17, 2009**

## INTRODUCTION / OVERVIEW

This report was commissioned by a committee of national experts in science and science education that has been convened by the National Research Council's Board on Science Education. The documents reviewed in this report include evaluations submitted from within most NOAA programs (NOAA Education Initiative --Science on a Sphere, 3; National Sea Grants, 3; National Marine Sanctuaries Program, 1; National Estuarine Research Reserves, 3; Ocean Exploration, 4; National Marine Fisheries Program , 1; and the Bay Watershed Education and Training Programs: Chesapeake Bay, California, and Hawaii), 1; National Ocean Sciences Bowl, 2. No evaluation reports were submitted by the Teacher at Sea program, Ocean Kiosk, NWS, Climate or JASON. Appendix A of this report lists the documents used for analysis in this report.

This report consists of the following sections: Purposes, Method of Analysis, Findings: Evaluation, Findings: Program, Reflections and Recommendations: Evaluation, and Reflections and Recommendations: Program.

## PURPOSES

The major purposes of this review of NOAA evaluation reports are to:

- Present a clear picture of the range and variety of evaluations conducted, including strengths and weaknesses of design, implementation, instrumentation, data analysis, interpretation of findings, recommendations and potential usefulness at the local, program, and national level; and
- Identify program strengths and highly regarded practices that emerge from the evaluations that may warrant sharing across programs or bringing to a larger scale.

## METHOD OF ANALYSIS

The following procedure was used in conducting this document analysis:

- Providing an initial, draft outline of the review to the Committee for approval
- Reading of all reports and general note-taking
- Identification of common data areas (e.g., evaluation questions addressed, evaluator recommendations) to be analyzed across all evaluation reports and construction of a data chart to use for note taking during the second review of the reports
- Completion of the data chart for each evaluation report

- Content analysis across all reports for each data area.

The analysis above was conducted with two separate lenses—the nature and quality of the evaluations and the strengths and highly regarded practices of the programs or projects being evaluated. These are presented in the two separate Findings sections below. The data chart used for note taking is attached in Appendix B.

Please note that quantifications provided within the Findings sections are simple counts rather than percentages, as the percentages can be somewhat misleading with these small numbers (18 reports). Also, the reader should be aware that in several cases there are multiple reports from one project or program with the same evaluator, so that the numbers within a positive or negative finding may be exaggerated.

## FINDINGS: Evaluation

This section of the review provides a brief description of the scope of the evaluation reports submitted and analyzed as well as how the 18 NOAA program evaluation reports addressed key elements of program evaluation. The evaluation reports provide a wide variety of programs, as well as evaluation approaches, size, and quality. The type of evaluator is very different across reports and, although no funding information was provided for analysis, it is readily apparent that the variation in funding for the evaluations was remarkable. Key elements addressed below following the description of the scope of the evaluation reports analyzed include: the evaluators, evaluation questions, evaluation designs, stakeholders as data sources, data collection strategies and instruments, quantitative data and analysis, qualitative data and analysis, limitations stated by evaluators, recommendations made by evaluators, reporting, consistency of evaluation plans with NOAA Evaluation Frameworks (Logic Models), notable evaluation strengths, and notable evaluation weaknesses.

### *Scope of Evaluation Reports Analyzed*

The documents reviewed in this report include evaluations submitted from within most, but not all, NOAA programs. In the section below, the scope of the evaluations used for this analysis is presented by program.

NOAA Education Initiative (NEI). Three evaluation reports were submitted related to this grant program that funds a range of projects such as museum exhibits, science lab course, teacher professional development and instructional materials, and a citizen weather monitoring program. The three reports were formative evaluations of pilot uses of the Science on a Sphere museum exhibit—one the initial pilot in MD in 2004, and the other two similar formative evaluations at the beginning and end of the exhibit period in MN in 2006. The same external evaluator conducted all three evaluations.

National Sea Grant Program. This program that consists of a national network of 30 university-based science education outreach programs that provide information to decision-makers and educators submitted three reports. One consisted of an internal online questionnaire of 46 members of the Sea Grant Network conducted in 2008 and

used as a tool to gather information about the programs and their needs, but not actually an evaluation. The other two reports provided evaluation information on two separate teacher learning projects: Teacher Education at Stone Laboratory (Ohio, undated) and the Aquatic Invaders in Maine (AIM) Teacher Workshop. The Stone Laboratory evaluation was conducted by a graduate research class (internal, because the members of the evaluation team had participated in the coursework themselves), and the AIM project evaluation was conducted by an external evaluator (graduate student).

National Marine Sanctuaries Program. This program that consists of 13 National Marine Sanctuaries that promote public understanding of national marine sanctuaries and the marine environment provided one evaluation report providing primarily description of the implementation of one teacher-student-scientist collaborative learning project, the Hawaii Field Study in 2005, with some formative, participant satisfaction information. It was conducted by an external evaluator.

National Estuarine Research Reserves System (NERRS). This program that includes 27 protected estuarine areas promotes estuary stewardship and provides a national coastal training program for coastal decision-makers as well as the national, interactive EstuaryLive program for students and teachers. NERRS submitted three reports. The first is the Inventory and Assessment of K-12 and Professional Teacher Development Programs in the National Estuarine Research Reserve System from 2003 that included questionnaire data from four different staff role groups at 24 NERR sites. In addition, this program provided evaluations from two consecutive years (2005 and 2006) of the national level EstuaryLive program. All three evaluations were conducted by the same external evaluator.

Office of Ocean Exploration. This program provides a range of K-12 education programs focused on using near real-time ocean discoveries to create excitement and enhance environmental literacy among teachers and students. The program submitted two evaluation reports. One, Evaluation of the Benefits to Scientists of Participation in Outreach and Education Projects Related to Ocean Explorations Expeditions (undated), was an internally-conducted survey of participating NOAA scientists from across a range of projects. Another report (undated) provided an internal evaluation of the Mountains in the Sea—Exploring the New England Seamount Chain project.

National Marine Fisheries Program (NMFP) This program submitted one evaluation report—an internal evaluation (2005) of the NOAA Science Camp in Seattle, WA.

Bay Watershed Education and Training (B-WET) Programs: Chesapeake Bay, California, and Hawaii. The B-WET program submitted one report—a large, external evaluation of the Chesapeake Bay area training program (DEL, MD, VA) that included teacher and student data from across many smaller projects within the area.

National Ocean Sciences Bowl This national program for high school students interested in ocean sciences provides mentoring and support of student teams that participate in academic competitions on the local, regional, and national levels. The two documents submitted for analysis included a summary of a national evaluation of the program in 2006 as well as a report of an ongoing longitudinal study: The Impact of the NOSB System on Participants' College and Career Choices in Science Disciplines (2007). Both evaluations were conducted by the same external evaluators.

National Ocean Service This line office submitted two evaluations of the NOAA WaterWays pilot ocean service project in 2007 and 2008. These evaluations were conducted by an evaluator using an integrated approach, also providing professional development within the project.

Teacher at Sea, Ocean Kiosk, NWS, Climate, JASON. No evaluation reports were submitted for these programs.

In summary, ten of the 18 reports in this collection evaluated a single project within a program, while seven provided program (cross-project) evaluations or surveys. One evaluation reported across projects within one of three regions of a program (B-WET). All 18 of the reports included some sort of formative evaluation information and seven also included a focus on impact. The reports include two surveys across programs, evaluations of one summer program for students only, three programs for teachers only, five programs for teachers and students, three programs that include live, online activities, one high school competition program, two research cruises, and three evaluations of a museum exhibit.

### *The Evaluators*

The perspective of the evaluator is an important component to consider. Generally, it is preferable to have an external evaluator to insure objectivity and quality of evaluation preparation. Internal evaluators are often used for cost or convenience reasons, and can be very effective if they take an objective approach and are well prepared in program evaluation. Over half (11) of the evaluators noted in the 18 reports appeared to be external, four internal, and three reports indicated other variations. The other variations included an evaluator for two reports who served within an integrated role as evaluator and also professional developer within the program. This was a unique approach in which the evaluator trained teachers in the use and development of concept maps and electronic portfolios to use as evaluation tools in the project. A third evaluation, because of lack of funding, was conducted by a university professor and members of a graduate course in educational research, all with previous experience with the project, either as an instructor or teacher participant in the program.

### *Evaluation Questions*

Evaluation questions are the cornerstone of any program evaluation. They are the key organizers of a strong evaluation, dictating the design of the study, the data collection strategies and instruments to be used, and the data analysis. The findings of the

evaluation provide answers to these questions and the basis for interpretation of findings and recommendations. Although actual questions are generally encouraged, it is common practice among many evaluators to take a shorthand approach by referring to program objectives or goals as the guides for the evaluation. In the case of this collection of 18 reports: seven provided specific evaluation questions; six provided clearly implied questions (referring to detailed program objectives, etc.); four included somewhat or barely implied questions; and one provided neither questions nor implied questions.

The foci of the stated or implied evaluation questions addressed a variety of topics, with the most frequently occurring ones being student learning/achievement (8 reports), student stewardship (6), and professional development design and strategies (5). A complete list follows in Table 1:

**Table 1: Foci of Evaluation Questions in 18 NOAA Program Evaluation Reports**

| **FOCI OF EVALUATION QUESTIONS (number of reports)** | |
|---|---|
| Student learning/achievement (8)<br>Student stewardship (6)<br>Student interest in science, careers (3)<br>Student satisfaction with activity (2)<br>Student engagement in learning (1)<br>Student sense of place (1)<br>Student leadership (1) | Teacher learning (4)<br>Teacher confidence in teaching ocean science (4)<br>Teacher satisfaction with PD (4)<br>Teacher implementation or intent to implement practices, use materials (4)<br>Teacher technology skills (2)<br>Teacher stewardship (1)<br>Teacher sense of place (1) |
| Scientist satisfaction with activities (2)<br>Scientist learning (1)<br><br>Museum visitor understanding/learning (3)<br>Museum visitor satisfaction (3)<br>Museum visitor suggestions for improvement (3) | Professional Development provided, strategies used, program design (5)<br>Professional Development evaluation used (2)<br><br>Program work environment (1) |

As was pointed out by one of the committee members, a previous NOAA strategic plan emphasized the importance of getting NOAA science in use through the NOAA education programs. As evident above, none of the program evaluations reviewed indicated an evaluation question or program objective directly focused on use of NOAA science. It should be noted, however, that the use of NOAA scientific research and

researchers was an underlying piece of most of the programs evaluated. Specifically, 15 of the 18 reports indicated in some way use of NOAA science and/or researchers as part of the program's work. Fourteen of the 18 reports provided teacher, student, or scientist satisfaction data concerning provision of the NOAA science research or data, or involvement of scientists in learning activities. Seven of the reports noted measures (self-report, tests, student presentations, or use of NOAA data) of teacher or student learning of NOAA-provided science content. Three of the reports gave no indications of a focus on using NOAA science, although one of these did provide a recommendation to develop a program using a system's research information.

### *Evaluation Designs*

Evaluation designs can vary widely, depending on the formative or summative purposes and the data available. All 18 reports had at least partially a basic descriptive design. A third of the evaluations (6) incorporated pre- post- elements (questionnaires or assessments). Two of the evaluations used control groups, although one of these was of negligible meaning because of extremely small numbers (intervention group = 7 teachers, control group = 4). Based on the reading of the reports there is no evidence of any sharing across programs of evaluators, evaluation designs, or metrics.

### *Stakeholders as Data Sources*

Particularly for formative evaluation of programs involving multiple stakeholders, it is generally recommended to gather data from as many of these groups as feasible. Thirteen reports provided data gathered from teachers, 13 from students, eight from program staff, coaches, or planning groups, four from scientists, three from the general public, and two from program interns. Nine of the evaluations used data from three or more stakeholder groups, four from two groups, and five from one stakeholder group only.

The numbers of stakeholders used as data sources varied widely in the 18 reports. Six evaluations reported 50 or fewer stakeholders; two provided data from between 51 and 100 stakeholders; six analyzed data from between 101 and 300 stakeholders; and four provided data from greater than 300. The variation in numbers, of course, is not surprising given the variation in size and scope of the 18 evaluations.

### *Data Collection Strategies and Instruments*

Another general rule of thumb for a strong evaluation is to use a variety of data collection strategies to maximize triangulation of findings and so that the methods match the questions to be answered by the evaluations. Seventeen of the evaluations used at least one questionnaire of program stakeholders (teachers, 8; students, 7; program staff, 4; scientists, 2), while 11 used interviews. Ten of the reports indicated observations of some sort (professional development, 2; student classrooms, 3; student or student and teacher activity, 2; museum, 3; other general site visits, 2). Ten also used some sort of learning assessment, aside from self-reported learning (teachers, 1; students, 7; general public, 3).

Evaluation instruments such as questionnaires, tests, and interview and observation protocols are often custom-made, as many were in the reports reviewed. This customization, of course, is helpful in gathering data that are specific to unique programs. On the other hand, these instruments lack validity and reliability data. One evaluation report noted a previously used questionnaire measuring student attitudes in the area of environmental stewardship that was previously used and tested for validity and reliability. Other instruments were custom-created and generally not pilot-tested, with only one report mentioning any piloting of instruments. Some of the evaluators did not include copies of instruments used in the appendix of their reports, but many questionnaire items could be reviewed within the results sections of the reports. Most of the questionnaires were reasonably well constructed, although some had minor problems such as poor match to the scale, confusing language, and lack of adequate open-ended items. One of the interview protocols tended to ask very narrow, closed questions that yielded very little useful information. Other instruments used included observation forms, student tests, and rubrics for assessment of student projects and presentations.

### *Quantitative Data and Analysis*

Quantitative data used in these evaluation reports were primarily from questionnaires—especially scaled attitudinal items and self-assessments of learning. Assessment of learning data included state science test scores in one evaluation, analysis of student products (e.g., rubrics for Power Point presentations) in three, and assessments embedded in questionnaires for teachers or students in four. Museum observations in three reports also provided quantitative data (time spent at the exhibit).

The majority of the analyses were basic descriptive statistics, including counts, percents, means, medians, and standard deviations. Informative graphs and tables were provided in many reports. Pre- post-data were analyzed in a few evaluations to determine changes in attitudes and knowledge, using non-parametric tests (sign test, Wilcoxen) as well as t tests. Chi Square was used in one evaluation to determine if the proportion of students choosing to attend certain universities was greater than expected by chance. Two studies conducted factor analysis and cluster analysis--the factor analysis to determine groupings of preferred professional development components, and the cluster analysis to determine groupings of surveyed projects that had similar needs or preferences. Linear regression and hierarchical linear modeling were also used in two separate reports to determine if the segment of an online program was a predictive factor of student scores in one evaluation and to assess the impact of a certain project approach on student environmental stewardship in another.

### *Qualitative Data and Analysis*

Sixteen of the 18 reports included some sort of qualitative data, with 11 providing analysis of open-ended questionnaire items, 11 providing interview data, and one with open-ended observation data. One report had no qualitative aspects and the other was unclear about qualitative data gathered. The analysis performed on the qualitative data was extremely variable in quality. Eleven of the reports did identify themes or

categories of responses from the interviews and open-ended questionnaire items. Four of those provided illustrative quotations with the categories/themes, and seven of these reports also quantified the responses within categories. Within six reports at least some items were not analyzed whatsoever beyond a list of comments or quotations. Four reports provided narrative descriptions of observations.

### *Limitations Stated by Evaluators*

It is a generally accepted research and evaluation practice to state directly any limitations to the study. Ten of the reviewed reports did not state any limitations to the evaluations. Six evaluations noted sampling concerns: two that random sampling of school classes was not used; one that self-selection of participants into the study creates a potential (positive) bias; six indicating the limitations of small samples, noting low response rates, lack of national advertising for teachers who might tune into live broadcasts and respond to online questionnaires, a need to expand a database with potential participants, and challenges of maintaining participation of subjects in a longitudinal study. One evaluation noted limited access to student test scores and another noted flaws in questionnaires (not including open-ended items and a set of items that was confusing to respondents).

### *Recommendations Made by Evaluators*

Evaluation reports are expected to include some interpretation of the findings along with program recommendations and often recommendations for further study. Eleven of the 18 evaluation reports provided at least one data-based recommendation to the program being evaluated. The recommendations or presentation of recommendations within three of these reports were somewhat problematic:

- Recommendations in one report were not clearly connected to the data that were analyzed and reported, but were apparently based on (unreported) staff feedback.
- In another report that provided extensive, high-level quantitative analysis that made up most of the results, the recommendations, though reasonable, were not connected to the statistical analysis, but rather to unreported or analyzed data from interviews and site visit conversations.
- One report gave only one very small substantive recommendation, ignoring data (suggestions for improvement) from participants.

### *Reporting*

Reporting, of course, is a very important aspect of any evaluation. The 18 reports analyzed varied greatly, especially in completeness. Data from analysis of completeness of reporting sheds some light on the quality of the evaluation reports, however, what is unknown by doing a simple document analysis is the reasons for certain missing sections, for example, funder needs or requests (often not expecting detailed descriptions of programs, literature reviews, or statements of limitations). Table 2 provides information concerning sections (or information) missing from the reports. Three of the evaluation reports had no missing sections. The most frequent missing sections in other reports are literature review (11) and limitations (10).

Perhaps more serious and surprising, however , is the fact that seven reports provided no actual description of the program(s) being evaluated, five reports had no stated or clearly implied evaluation questions, eight had no discussion or interpretation of findings, and seven had no recommendations.

**Table 2: Missing Evaluation Sections or Information, n = 18 reports**

| MISSING EVALUATION SECTIONS | NUMBER OF REPORTS |
|---|---|
| Description of program | 7 |
| Literature review | 11 |
| Evaluation questions | 5 |
| Methods description | 0 |
| Results description | 0 |
| Limitations | 10 |
| Discussion or interpretation of results | 8 |
| Recommendations | 7 |
| Protocols and questionnaires included in appendix | 5 |

### *Consistency of Evaluation Plans with NOAA Evaluation Frameworks (Logic Model)*

One of the background documents provided for this analysis is the National Academies Education Evaluation Report submitted by the NOAA Office of Education. This document provides a brief description and graphic of the *Targeting Outcomes of Programs* (TOP) evaluation and planning logic model that was adopted by the NOAA Education Council in 2007 to serve as a common framework for evaluation of NOAA education programs. The description of the levels of assessment (or areas to be evaluated) from this document is presented below.

"The seven assessment levels are briefly defined as follows:

- **Resources** level explains the scope of the programming effort in terms of dollars expended and staff time used.
- Progress documented at the **Activities and Participation** levels generally is referred to as outputs. It indicates the volume of work accomplished and is evidence of program implementation.
- **Reactions**, is evidence of participants' immediate satisfaction.
- Intermediate outcomes at the **KASA (knowledge, attitude, skills and aspirations)** level focus on knowledge gained/retained, attitudes changed, skills acquired, and aspirations changed.
- Intermediate outcomes at the **Practices/Behavioral** level focus on the extent to which best management practices are implemented by program participants and others whom they may influence. These intermediate outcomes can be measured months or years after program implementation.
- Intermediate outcomes lead to longer term **social, economic, and environmental (SEE)** changes. Identifying outcomes at the SEE level for individuals and localities may occur fairly quickly but state, regional, or national outcomes may take years to assess and be very expensive."

(National Academies Education Evaluation Report submitted by the NOAA Office of Education, pp. 1-2)

This logic model was used to determine to what extent the 18 NOAA evaluation reports submitted for analysis addressed the seven levels of assessment as encouraged by this guidance document. (It should be noted that many of the evaluation reports submitted were dated prior to this guidance document. Also, it is unknown to what extent this logic model was shared across NOAA programs.)

Seventeen of the 18 reports provided some data concerning Program Activities and Participation, although this information was often spotty in nature. All 18 reports provided information on the Reaction level. Sixteen reports provided data concerning intermediate outcomes in the area of Knowledge, Attitude, Skills, and Aspirations. Nine included information concerning intermediate outcomes in the area of Practices or Behaviors. None of the reports reported information concerning broader Social, Economic, or Environmental changes.

### *Notable Evaluation Strengths*

In reviewing the evaluation reports, particular strengths that stood out for each were noted. It is important to realize, of course, that this is one person's take on these strengths, but nevertheless, perhaps useful for the Committee's broader view. Some reports certainly had more strengths than others, and some were very strong in certain areas and not at all in others.

The primary areas of strength noted in some of the evaluation reports were:

- Reporting from multiple sources, e.g., teachers, students, staff, interns, scientists
- Providing useful formative data and recommendations for programs/projects
- Providing informative program impact data
- Effectively using and presenting descriptive statistical analyses
- Effectively using and presenting inferential statistics
- Rigorous, artful, and informative presentation of qualitative findings
- Providing a particularly effective balance of quantitative and qualitative data
- Including insightful literature reviews that were used to analyze program design and interpret findings concerning program implementation
- Providing particularly clear, well-written overall reports.

### *Notable Evaluation Weaknesses*

Likewise, in reviewing each evaluation report, particularly serious weaknesses were also noted in some reports. The following summarizes these weaknesses:

<u>Clarity and focus</u>

- Lack of description of the program being evaluated
- Lack of evaluation questions to focus the report
- Lack of conceptual framework needed in some cases

<u>Methodology and instrumentation</u>

- Missing detail on methods, e.g., sample selection, questionnaire piloting and administration
- Some over-dependence on self-report

- Over-dependence on either qualitative or quantitative data
- Some questionnaires and interviews rather poorly constructed
- Some questionnaire items designed more for the data analysis than for understandability for the respondents or for finding out what the respondents actually think (and likely were never pilot tested)
- Comparison study using a control group of students has very little information on the implementation of the program except for the number of hours spent which varies a great deal

Data analysis and presentation of results

- Some data collected, but unreported (interviews, site visits) or unanalyzed (analysis across classroom observations)
- Poor or no analysis of qualitative data, e.g., giving a simple list of comments
- Poor presentation of results of basic descriptive statistical analyses, e.g., unclear graphs, missing n's
- Poor use of data from multiple sources so that even though many different role groups were surveyed, their roles and specific concerns were not made clear
- Complex statistical analysis that may be more than what is needed for the purposes of the evaluation and answering the evaluation questions
- Some statistical analysis apparently done mostly "for the sake of doing statistics" since these findings were ignored and not used to inform recommendations
- Extensive statistical analysis using national data across programs, yet presented pretty much in isolation, with no evaluation questions, no discussion, and no recommendations

Interpretation of results and recommendations

- No reflection, interpretation, or discussion of findings in ways that might help programs improve
- Use of unanalyzed or unreported qualitative data to make recommendations
- Evaluator so focused on the conceptual framework of the program that he/she neglects to bring forth what was actually asked of respondents and what they had to say about the program, leading to an artificial analysis of data and weak recommendations.
- No recommendations presented at all.

The following section of this report presents the findings of the cross-report analysis based on program. Reflections and recommendations concerning evaluation and program are provided later in the report.

## FINDINGS: Program

The focus of this section of this report shifts from the evaluations conducted to the programs evaluated. It is very important for the Committee members using this report to realize that this is in no way a What Works Clearinghouse type of meta-analysis, identifying strong programs only if their program results were based on stringent research requirements such as experimental designs. None of the NOAA evaluation reports submitted meet those requirements. In addition, it is clear that some apparently strong programs had rather flawed evaluations. Rather, this summary of the

notable program findings is based on the individual reports' results taken at face value as presented and based on evaluator interpretations of findings (when those are given). These evaluations are based on a great deal of self-report (some of it well documented and some not) from teachers, scientists, students, NOAA program staff members, and museum visitors. This summary is also based on findings from a variety of measures of teacher and student learning and analysis of student products. It is also important to note that some reports included minimal or no substantive description of the actual programs that they evaluated. In this section the results are presented in the following areas: teacher satisfaction and identification of program strengths; impact of programs on teachers; barriers for teachers; teacher suggestions for program improvements; program challenges; student satisfaction; student outcomes; scientist satisfaction, concerns, and impact; museum exhibit satisfaction and needs for improvement; program evaluator recommendations; and highly regarded practices.

### *Teacher Satisfaction and Identification of Program Strengths*

All of the program evaluation reports that included teacher data noted high levels of satisfaction from teacher participants. Teachers consistently noted appreciation for NOAA program learning opportunities. The areas of satisfaction noted in different reports include:

- Hands-on learning in natural environments (outdoors)
- Working with NOAA scientists, especially in natural environments such as ocean cruises, estuary exploration
- Seeing in person or in real time online how science research really actually takes place
- Modeling of scientific research
- Combinations of classroom learning and fieldwork
- Time to practice new skills
- Time to plan for integration of new learning and skills into their own classroom lessons or curriculum
- Time and activities that build collegial relationships and collaboration, networking among teachers
- Teachers designing their own lessons/activities/units based on PD learning, concept maps, and identification of student misconceptions or gaps in understanding
- Teachers developing and implementing problem based learning units
- Teachers and students accessing and using online NOAA data
- Interactive social groupings of teachers, students, and scientists
- Using photography as a tool in science research
- Relationship building among students, teachers, scientists, and program staff
- Immersion in the local environment and developing a "sense of place"
- Teachers, scientists, and program staff who serve as mentors and role models

### *Impact of Programs on Teachers*

The purposes or desired outcomes of the programs being evaluated were not always clearly stated. Some of the impacts on teachers noted by evaluations, however, are:

- Increase in teacher confidence to teach about NOAA-related topics
- Development of related lessons, activities, or units to use with their students
- Intent to continue teaching about NOAA-related topics or environmental issues
- Increased understanding of NOAA-related scientific content
- Better understanding of how scientific research actually happens
- Productive relationships with other teachers, scientists, and students

### *Barriers for Teachers*

Some of the barriers reportedly facing teachers who wished to implement NOAA program learning and processes include:

- Lack of clear connections with required standards
- Insufficient flexibility in curriculum
- Insufficient funding for equipment, field trips
- Insufficient time to collaborate with other teachers
- Technology problems in accessing online programs and data

### *Teacher Suggestions for Program Improvements*

Numerous reports noted teacher suggestions for program improvements. These include the following:

- Better connections of program work with standards used in schools
- Multi-day programs
- More time for teachers to work together on lessons and curriculum work
- More follow-up support for teachers
- Support for teachers to maintain collaboration with other teachers in different schools or locations
- More evaluation and assessment tools online
- More NOAA scientists involved
- More examples of lessons for teachers
- More preparation for teachers before cruises and real-time online activities
- Less lecturing by scientists on live online programs
- Better audio and visual quality on live online programs
- More and better user-friendly materials for teachers and students

### *Program Challenges*

A few challenges facing NOAA programs were mentioned, mostly from program staff:

- Increasing difficulty in getting students and teachers to visit reserves
- Little program evaluation
- Mixed views across similar programs concerning usefulness of developing national programs and collaborating with other programs
- Communicating broadly about live online programs
- Inadequate preparation by teachers and students when they tune into live online programs

### *Student Satisfaction*

In a few reports data were collected from students concerning their satisfaction with the programs that had touched them.

- Students and their teachers report that students are excited by live online programs when they are able to see their peers working with scientists to conduct real ocean or estuary research.
- Students are excited when they can ask questions of the researchers—they appreciate the interactivity.
- Students enjoy and are engaged by live broadcasts online that allow them to see the diversity of estuary life forms and compare estuaries in distant places.
- Student enjoy learning outdoors.
- Students learn from and enjoy collecting and analyzing data.
- Students enjoy learning from educational videos, internet information, and hands-on activities or labs.
- High school students involved in ocean sciences competitions value greatly their team mentors.

### *Student Outcomes*

A number of important student outcomes were determined, some by pre- post- self report of teachers and students and pre-post tests of students:

- Increased interest in studying the oceans
- Increased knowledge of the oceans, weather, climate, estuaries, invasive species, monitoring of growth in habitats
- Increased interest in science in general
- Students see how science works, especially the roles of curiosity and risk taking
- Ability to use online NOAA data and NOAA internet sites to learn about oceans
- Ability to develop a presentation for NOAA scientists
- Awareness of human activities that may impair estuaries
- High school students influenced by NOSB in career choice, college major, and overall interest in oceans
- NOSB high school students have maintained relationships with team-mates and coaches
- Influence on high school student environmental stewardship
- Leadership skills, including confidence, planning, working relationships, and ethics among high school students participating in academic competitions.

### *Scientist Satisfaction, Concerns, and Impact*

One study surveyed scientists and found :

- General satisfaction with many aspects of their work with teachers and students, seeing greater understanding of habitats and ecosystems studies
- Need for more feedback for scientists on their work in educational outreach
- Concerns about lack of pay, time, and recognition for scientists' outreach work
- More people showing interest in scientists' research.

### *Museum Exhibit Satisfaction and Needs for Improvement*

Three reports provided formative evaluation of the Science on a Sphere exhibit. Findings included:

- Visitors enjoyed the experience of the SOS, including the live presentations.
- Visitors were most interested in the technology behind the exhibit.
- Visitors made suggestions for improvement in the presentation, including labels, more audio, providing different views.
- SOS effective in demonstrating a "range of levels of scientific ideas and facts from big ideas to facts".
- Visitors stay longer at this exhibit than others.
- Challenges: how to make the exhibit more self-standing and interactive and how to get beyond the wow factor.
- Overall purpose of the exhibit is not that clear to visitors.

### *Program Evaluator Recommendations*

Eleven of the 18 reports reviewed provided recommendations for programs and/or further study. The most relevant ones included:

Professional Development Pre-Work

- Establish better communication to standards used in schools
- Provide more information and preparation for teachers and students before live broadcast programs
- Provide more opportunities for interactivity between teachers and scientists before programs, e.g., teachers providing their standards and plans for lessons before the cruise or other activity
- More scholarships for teachers

Professional Development Design and Implementation

- Multi-day programs
- Professional development for teams of teachers
- More time for teachers to work on their own curricula and standards and how the program fits into their work
- Provide more materials and example lessons for teachers
- Provide evaluation and assessment tools online
- Increase the number of scientists involved
- Provide more technical support for live broadcast programs
- Provide more interactivity in live broadcast programs
- More time for curriculum development and group work for teachers

Professional Development Follow-Up

- Support for teacher collaboration
- Follow-up support for teachers
- Support school to school collaboration

Student Camp

- Revision of environmental scenario required of student campers

- Reduction in length of camp

High School Student Competition

- Better communication of career information to high school students participating in NOSB
- Enhancement and marketing of leadership components of NOSB
- Professional development for NOSB mentors
- Expanded communication with parents about NOSB

Future Evaluation

- Research other measures of student achievement (beyond state science test scores)
- Improvement and expansion of student database to support long-term research

### *Highly Regarded Practices*

Five general practices or approaches emerge from the NOAA program evaluations and cross-program surveys reviewed for quality and potential scalability. These practices appear in a mix of different programs to greater or lesser degrees and are supported by evidence of participant and staff satisfaction as well as impact on teachers, students, and scientists. In addition, although this current review did not include a literature review, it should be noted that basic aspects of the following approaches have some degree of a positive research base. The five practices are presented below with examples from some of the program evaluations reviewed.

Teacher (and Student) Learning in Collaboration with Scientists and NOAA Program Staff

NOAA scientists and staff connect with educational outreach programs in a variety of ways across programs, yielding a great deal of appreciation, inspiration, and learning on the part of program participants.

- *Sharing of research through lectures and other methods*. NOAA Science Camp, Aquatic Invaders in Maine, Mountains in the Sea—Exploring the New England Seamount Chain, Chesapeake Bay Watershed Education and Training Program: Meaningful Watershed Educational Experiences
- *Exploration of ocean environment*. National Marine Sanctuary Program Hawaii, Mountains in the Sea—Exploring the New England Seamount Chain
- *Demonstration of research fieldwork.* NERR System Programs, Chesapeake Bay Watershed Education and Training Program: Meaningful Watershed Educational Experiences
- *Interacting with summer camp attendees.* NOAA Science Camp
- *Interacting with students online about their research and answering questions students pose.* Mountains in the Sea—Exploring the New England Seamount Chain
- *Reviewing /assessing student presentations on problem based learning projects.* NOAA Waterways Ocean Service Pilot Project

- *Mentoring competing teams of high school students of ocean science.* National Ocean Sciences Bowl
- *Providing narrative explanations to accompany museum exhibit.* Science on a Sphere exhibits in Maryland and Minnesota

Teacher Professional Development Through Which Teachers Integrate Their Learning Into Lessons and Problem Based Learning Activities for Students Often Using NOAA Data Available Online

Some of the NOAA programs provide time, support, and materials for developing and integrating teacher content and process learning from NOAA into their teaching, developing lessons and problem based learning activities for students. This approach builds in teacher implementation, follow-up support, and engagement of students usually in team efforts and sometimes directly related to initial student questions.

- Chesapeake Bay Watershed Education and Training (B-WET) Program: Meaningful Watershed Educational Experiences (MWEE)
  *"A MWEE integrates field experiences in the Chesapeake Bay watershed with multi-disciplinary classroom activities and instruction…. Students then share their discoveries about the watershed with local schools and communities, both orally and in written form. MWEE's:*
  - *Are investigative or project-oriented,*
  - *Are integrated within the instructional program,*
  - *Involve preparation, action, and reflection,*
  - *Reveal the watershed as a system, and*
  - *Are integrated into a significant amount of instructional time, ideally a school year." (p. 5, Evaluation of National Oceanic and Atmospheric Administration Chesapeake Bay Watershed Education and Training Program Meaningful Watershed Educational Experiences)*
- Aquatic Invaders in Maine
- Mountains in the Sea—Exploring the New England Seamount Chain
- Estuary Live, 2005, 2006

Research Cruises or Other Explorations Using Online Broadcasts and Data

Research cruises and similar explorations provide direct experiential learning for onboard teachers and students, but potentially more importantly, some connect with classrooms of teachers and students across the nation, providing modeling of ocean and estuary exploration, research, and monitoring, as well as opportunities for interactive questions and answers between students and scientists and for students to see their peers engaging in real scientific research activities.

- EstuaryLive, 2005, 2006
- Mountains in the Sea—Exploring the New England Seamount Chain
- National Marine Sanctuary Program Hawaii (no online component)

High School Programs for Students Interested in the Ocean Sciences That Combine Competition, Collaboration, and Mentoring

The National Ocean Sciences Bowl is a national program for high school students that supports school teams in an academic competition at regional and national levels, providing coaching by teachers and scientists. The program is designed for high achieving students and promotes intensive learning, leadership, competition, and collaboration.

Field Learning in the Outdoors

Many of the NOAA programs involve teachers and/or students in outdoor learning activities that are deemed to be essential to the learning and especially to development of qualities of stewardship.

- NOAA Science Camp
- Chesapeake Bay Watershed Education and Training Program: Meaningful Watershed Educational Experiences
- EstuaryLive, 2005, 2006
- Mountains in the Sea—Exploring the New England Seamount Chain
- Teacher Education at Stone Laboratory

## REFLECTIONS: Evaluation

NOAA has an extensive and ambitious mission, and the collection of evaluation reports and other documents represents a broad mix of NOAA educational programs, with a range of approaches, sizes, audiences, and goals. The evaluation reports reflect a similar variety in almost all aspects of evaluation, but particularly in evaluation purposes, methodologies, data analysis, interpretation of findings and recommendations—from a small, focused formative evaluation of a museum exhibit, including observations and interviews of exhibit visitors, to an evaluation of the impact of a large teacher professional development initiative involving extensive teacher and student pre/post questionnaires as well as analysis of student scores on state standardized science tests. The scope, quality, completeness, and potential usefulness of the evaluations in this collection are extremely varied.

The analysis of this set of reports identified many examples of strong evaluation elements, for example, clear evaluation questions with well-described methodologies appropriately matched to the questions and to the data available, effective presentation of qualitative and quantitative data, valuable interpretation of findings and offering of useful recommendations. Almost all of the reports had some redeeming qualities, however, very few reports provided a "complete package" of quality or usefulness. An accounting of many of the rather glaring weaknesses of the evaluations is located on pp. 7-8 of this report, including particularly lack of evaluation questions to guide the study, poorly constructed data collection instruments, poor qualitative data analysis, extensive quantitative analysis that was not used to address evaluation questions or to create useful recommendations.

NOAA is complex, with a multiplicity of goals and priorities and programs. Given these expansive goals and variety of approaches, there needs to be a substantive plan for

evaluation that provides essential formative evaluation for new and pilot programs as well as impact evaluation for more long-running, high-visibility, broadly implemented programs. High expectations and guidelines for program evaluation need to be communicated broadly and continuously at the national and program levels. This certainly will require development, documentation, and training. But more than this, it will require building a culture of evaluation that builds not only expertise and understanding of evaluation but also ownership and leadership. Evaluations need to focus clearly on program goals and support program leaders in achieving their goals to a maximum level. Program leaders need to understand the value of evaluation in helping them to clarify goals, provide useful feedback on program, and document impact. There also need to be adequate budget allocations for evaluation—otherwise damaging shortcuts are taken.

A certain level of consistency across program evaluations is to be desired, however, considerable cross-program thought and discussion will be needed to achieve this. Whereas it is desirable to expect and require consistent elements and levels of quality from all evaluations, the size, scope, design, and methods used will vary greatly based on the purpose of the given evaluation and the purpose of the actual programs. In addition, careful thought and in-depth discussion between program leaders and evaluators need to go into decisions about measuring outcomes of NOAA programs. Given the effort, time, and money that go into measuring and reporting outcomes, it is essential that the selected measurable outcomes are indeed reasonable to expect. For example, what can be reasonably expected (and then measured) in terms of student outcomes based on a week-long teacher professional development program? What kind of measurable impact on public environmental stewardship can be expected from a museum exhibit that people explore for less than five minutes? This is not to disparage programs such as these or their potential impact, it is just a call to carefully match the kind of evaluation work and the outcome measures selected to the actual purpose, approach, and reach of a given program. This careful matching of the evaluation to the program is likely much more important than consistency of methods, designs, or measures used. And, of course, consistency across program evaluations could certainly be encouraged, in terms of essentials of good evaluation, such as clear evaluation questions, appropriate use of quantitative and qualitative data, and program recommendations that relate to the findings of the study.

It is important to remember that program evaluation is most of all about learning—learning what works well in which contexts with which stakeholder groups, learning how programs can be improved, and learning about the multitude of intended and unintended outcomes of innovative initiatives. Evaluation is also about learning about effective approaches and instrumentation for answering critical evaluation questions. Evaluation can be an enormously effective tool for cross-program learning and maintaining focus on a common mission. However, this does not happen easily or by simply requiring evaluations or even by distributing well-constructed evaluation expectations. It requires a great deal of ongoing cross-program talk involving program staff and evaluators. NOAA's apparent strength of creating and supporting different

and innovative programs can certainly be maximized by sharing and really using different program perspectives and the findings of the evaluators about program elements as well as evaluation approaches.

## REFLECTIONS: Program

Based on what can be gleaned from the evaluation reports used for this study, NOAA provides a wide variety of well-designed and implemented programs for teachers, students, and the general public, often involving or incorporating NOAA scientists, NOAA research and data, and, of course, natural resource areas nurtured and protected by NOAA. The greatest challenge for NOAA appears to be to establish a clear purpose for each project and then to design it to meet that purpose and to implement it as designed, with high quality, and with a clear plan for increasing reach and impact. As noted, many positive strategies are present in many of these programs, especially the following five:

- Teacher (and student) learning in collaboration with scientists and NOAA program staff
- Teacher professional development through which teachers integrate their learning into lessons and problem based learning activities for students often using NOAA data available online
- Research cruises or other explorations using online broadcasts and data
- High school programs for students interested in the ocean sciences that combine competition, collaboration, and mentoring
- Field learning in the outdoors.

But any of these components alone is not adequate, particularly if certain ones are not well implemented. For example, involving scientists in explaining their research to teachers and students is a positive element—unless, the presentation is a lengthy, boring, "talking head" lecture. As also noted about the evaluation reports, almost all of the projects seemed to have some truly positive aspects, however, few seemed to provide a "complete package" of high quality approaches. There appear to be significant missing pieces within many of the programs, e.g., programs with apparently strong teacher professional development, but with no follow-up support during the school year; a well-designed learning experience for a small number of teachers and students, with again, no follow-up or plan for scaling up; live, on-line programs that do not provide adequate pre-activity preparation for participants. It seems there needs to be greater attention paid to important work that goes before a learning activity as well as the work that occurs afterwards. These "fuller packages" hold greater promise for impact on learning. This is also where the role of formative evaluation becomes so important—enabling projects to build on their strengths and add or improve weaker parts. Project staff need to demonstrate that they hear the feedback being received and use it to make their projects better meet the needs of participants. Pilot programs should not be funded without a plan for scaling; nor should they be scaled up without addressing issues raised by formative evaluation.

Communication appears to be a concern for some projects. Some programs link closely to schools, e.g., through professional development of teachers, direct involvement of students in programming and exploration of natural resources, lesson planning, and such. However, from the limited information available in the evaluation reports, it seems that communication and collaboration may be spotty or inconsistent, e.g., with a few teachers in a few schools, or with just a few schools in a district, or with the schools, and not the district staff. These limited approaches are likely to have limited impact. Communication was also mentioned as a challenge for one on-line live broadcast project, in that very few educators across the country knew about the opportunity for their teachers and students to participate. NOAA may want to encourage project staff to think carefully about potential communication challenges and effective communication strategies before a project begins.

Just a couple of the reports noted collaborations with other institutions such as National Geographic and NASA. These NOAA projects seemed to gain significantly from the collaborations. Such partnerships, when linked with careful attention to purpose and learning, may increase access to scientific data and scientists, and, of course, pooling of resources and funding.

The challenge of the committee convened by the NRC is to advise NOAA wisely concerning prioritizing programs and program components, when there are obviously many positive elements occurring. NOAA will be responsible for effective communication of expectations for programs in the future to receive funding and other support. As noted previously concerning evaluation, for program purposes as well, there is a need for the communication of high expectations and explicit guidelines for projects to be funded or otherwise supported. Program creation and development will also be greatly enhanced by ongoing cross-program conversations among staff and evaluators.

The analysis from this study of the evaluation reports points to some basic guidelines or filters for prioritization, planning, implementation, and evaluation of NOAA educational outreach projects:

- ***Clarity of purpose and desired outcomes*** from design through implementation and evaluation
- ***Maximum reach of identified audiences or participants***
- ***Incorporation and maximization of use of some of the five highly regarded practices*** or others that are piloted
- ***Plans for use of formative evaluation*** for improvement of projects
- ***Pilot projects that include clear plans for scaling up***
- ***Inclusion of explicit communication plans***
- ***Cross-institutional collaborations*** that complement and enhance NOAA contributions

- ***Projects that provide a "complete package"*** (purpose, activities, reach or scale, communication, formative evaluation, collaborations) in terms of design

## RECOMMENDATIONS: Evaluation and Program

Given the expansive goals and variety of approaches of NOAA's educational programs, NOAA and the NRC Committee reviewing these programs, while understanding the current program priorities and pressures, including budget constraints, may want to consider the following recommendations:

- ***Develop a strong and strategic plan for program development.*** What are the "most essential" purposes and audiences of NOAA's educational programs? What are the combinations of strategies that NOAA might try to more fully develop and improve given those essential purposes and audiences? Can certain successful programs such as the NOSB high school competition be used to reach different participants, e.g., students who may not currently be high achieving? What specific strategies are needed to build and sustain relationships with schools and school districts in ways that NOAA science and outreach can maximally contribute to teacher and student learning, rather than remaining in the fringes of enrichment activities? What other institutional collaborations might most effectively complement the strengths of NOAA? How can museum exhibits be designed for maximum reach and impact within the overall missions of the museums?
- ***Develop a strong and strategic plan and both formative and impact evaluation.*** What is the balance of formative and impact evaluation that is needed in the years to come? For example, which programs, particularly new ones or pilots, will need strong formative evaluation? Which longer-running and perhaps more broadly implemented programs warrant extensive (and often more costly) impact evaluations? Are there essential focus areas that need to be addressed by all program evaluations? What are the benefits and drawbacks of internal vs. external evaluators? How can program staff be trained to conduct ongoing self-study evaluations to supplement, but not replace, external studies?
- ***Provide frequent in-person cross-program forums*** for discussion and sharing of program strengths and evaluation findings and approaches. These forums for NOAA program staff and their evaluators would also provide opportunities for:
    - Introduction of new programmatic strategies
    - Provision of training/guidance concerning essential elements of high quality formative and outcome evaluation and effective collaborative evaluation planning and use of findings
    - On-going support and mentoring of program leaders in overseeing projects and evaluation efforts
    - Development of user-friendly materials, including simple checklists for essential components of project design as well as evaluation planning and reporting

- ***Approach budget decision-making concerning program and evaluation work*** based on cross-program discussions among program staff, with input from evaluators, and taking into account essential purposes of NOAA programming and evaluation as well as external funding and other pressures.

## APPENDICES

- **APPENDIX A: List and Characteristics of Reports Analyzed in This Report**
- **APPENDIX B: Cross-Report Analysis Note Taking Sheet**

## APPENDIX A: List and Characteristics of Evaluation Reports Analyzed in This Report

| PROGRAMS AND EVALUATION REPORTS | SINGLE PROJECT | CROSS PROJECT (PROGRAM) | FORMATIVE | IMPACT | TOTAL N | EVALUATOR* |
|---|---|---|---|---|---|---|
| **NOAA Education Initiative (NEI)** | | | | | | |
| Science on a Sphere (SOS) Front-End Evaluation Report 2004—MD Science Center | X (Pilot) | | X | | Approx. 400 | External A |
| SOS Front-End Evaluation Report 2006—Science Museum of MN | X | | X | | 81 | External A |
| SOS Formative Report 2006—Science Museum of MN | X | | X | | 50 | External A |
| **Bay Watershed Education and Training (B-WET) Programs: Chesapeake Bay, California, and Hawaii** | | | | | | |
| Chesapeake Bay Watershed Education and Training Program Meaningful Watershed Educational Experiences--2007 | | Cross-project evaluation of one of the three B-WET programs | X | X | 2828 | External B |
| **National Sea Grant Program** | | | | | | |
| Summary of internal online survey of Sea Grant Network 2008 | | X | X | | 46 | Internal A |
| Teacher Education at Stone Laboratory: Program Description, Literature Setting, | | | | | | Internal B (graduate |

| PROGRAMS AND EVALUATION REPORTS | SINGLE PROJECT | CROSS PROJECT (PROGRAM) | FORMATIVE | IMPACT | TOTAL N | EVALUATOR* |
|---|---|---|---|---|---|---|
| and Impact on Educators | X | | X | | 77 | class) |
| Aquatic Invaders in Maine (AIM) Teacher Workshop Evaluation | X | | X | X | 116 | External C (graduate student) |
| **Office of Ocean Exploration** | | | | | | |
| Advancing Collaboration: Analysis of the Effectiveness of Research and Education Partnerships at Sea Using Real-Time Technology and Professional Development (undated) | X | | X | | 42 | Internal C |
| Evaluation of the Benefits to Scientists of Participation in Outreach and Education Projects Related to Ocean Explorations Expeditions (undated) | | X | X | | 19 | Internal D |
| **National Estuarine Research Reserves System NERRS** | | | | | | |
| Inventory and Assessment of K-12 and Professional Teacher Development Programs in the National Estuarine Research Reserve System 2003 | | X | | X | Approx. 96 | External D |
| Evaluation of EstuaryLive 2005 | | X | X | X | Approx. 400 | External D |

| PROGRAMS AND EVALUATION REPORTS | SINGLE PROJECT | CROSS PROJECT (PROGRAM) | FORMATIVE | IMPACT | TOTAL N | EVALUATOR* |
|---|---|---|---|---|---|---|
| Evaluation of EstuaryLive 2006 | | X | X | X | Approx. 415 | External D |
| **National Marine Sanctuaries Program (NMSP)** | | | | | | |
| Evaluation of National Marine Sanctuary Program Hawaii Field Study 2005 | X | | X | | 24 | External E |
| **National Marine Fisheries Program (NMFP)** | | | | | | |
| Evaluation of NOAA Science Camp Seattle WA 2007 | X | | X | | Approx. 50 | Internal E |
| **National Ocean Sciences Bowl (NOSB)** | | | | | | |
| Summary of Evaluation of the National Ocean Sciences Bowl 2006 | | X | X | | Approx. 500 | External F |
| National Ocean Sciences Bowl Longitudinal Study: The Impact of the NOSB System on Participants' College and Career Choices in Science Disciplines | | X | X | X | Approx. 100 | External F |
| **National Ocean Service (NOS)** | | | | | | |
| NOAA WaterWays 2007: Evaluation of the NOAA Ocean Service Pilot Project | X | | X | X | Approx. 20 | Integrated A |
| NOAA WaterWays 2008: | | | | | | |

| PROGRAMS AND EVALUATION REPORTS | SINGLE PROJECT | CROSS PROJECT (PROGRAM) | FORMATIVE | IMPACT | TOTAL N | EVALUATOR* |
|---|---|---|---|---|---|---|
| Evaluation of the NOAA Ocean Service Pilot Project | X | | X | X | Approx. 40 | Integrated A |
| **Teacher at Sea** | | | | | | |
| No reports received | | | | | | |
| **Ocean Kiosk** | | | | | | |
| No reports received | | | | | | |
| **NWS** | | | | | | |
| No reports received | | | | | | |
| **Climate** | | | | | | |
| No reports received | | | | | | |
| **JASON** | | | | | | |
| No reports received | | | | | | |

***Letters are used to identify different evaluators, e.g., external evaluators, A-F (six different evaluators or evaluator groups), internal evaluators, A-E (five), and integrated evaluator, A (one).**

## APPENDIX B: Cross-Report Analysis Note Taking Sheet

| EVALUATION FINDINGS | |
|---|---|
| **ASPECT OF EVALUATION** | **REPORT #** |
| Type of evaluator | |
| Program description | |
| Evaluation Questions Addressed | |
| Defined Measures | |
| Design | |
| Data Collection Strategies | |
| Data Collection Instruments | |
| Stakeholders as Sources | |
| Scope of Data Collection | |
| Quantitative Data and Analysis | |
| Qualitative Data and Analysis | |
| Appropriateness of All to Answer Evaluation Questions | |
| Consistency with NOAA Guidelines | |
| Report Organization and Completeness | |
| Literature Inclusion | |
| Limitations Stated | |
| Other Evaluation Challenges Noted | |
| Discussion/Reflections | |
| Recommendations (Based on findings?) | |
| General Strengths of Evaluation | |
| General Weaknesses of Evaluation | |
| **PROGRAM FINDINGS** | |
| **ASPECT OF PROGRAM** | **REPORT #** |
| Essential Findings, e.g., Satisfaction, Application of Learning, Student Outcomes | |
| Program Strengths or Highly Regarded Practices | |
| Program Challenges | |
| Evaluator Recommendations for Improvement | |