

Summary and Critique of Selected Evaluations of NASA Educational Programs
Frances Lawrenz
University of Minnesota

Executive Summary

This paper presents an evaluative summary of 10 evaluation reports about NASA educational programs. The evaluations were conducted of six different NASA programs by six different evaluation providers over the years from 2003-2006:

- GLOBE program by SRI in 2005 and 2006
- Explorer Schools by Center for Educational Technologies in 2003-6 and by Paragon TEC in 2006
- Sun-Earth Day 2006—Eclipse: In a different light program by the Technology for Learning Consortium, Inc.
- Earth KAM Program by EDC in 2006
- Digital Learning Network Evaluation Tool Development program by the Technology for Learning Consortium, Inc. in 2006
- AESP program by the Evaluation Center in 2004-6.

Each report is considered in some depth through an overview of what the evaluation included as well as a critique of the evaluation questions, methods and findings. A table outlining the titles of the evaluations and who conducted them, the date of the report, the evaluation questions, the evaluation design or methods and brief comments on the quality of each report is also provided. A wide range of evaluation questions were considered although generally the questions related to describing what was happening in the program and considering the program's impact. The evaluation questions tended to shift within and between evaluations. A number of different methods were used but most popular were surveys and case studies which are largely based on participant self-reported information. There were two quasi-experimental studies conducted, both for GLOBE, but these were very limited. Two other quasi-experimental studies were proposed, both for the Explorer schools program, one by Center for Educational Technologies and one by Paragon TEC but these were not actually implemented.

The critique focused on methodology (design, data collection methods, analysis and reporting.) The reports were all somewhat typical large scale evaluation reports seemingly intended for only internal use in that they were descriptive of program operations. The reports themselves were generally of acceptable, but varying, quality with appendices presenting detailed information and brief, but informative, extracts of the data provided in tables in the body of the report. The designs were mostly retrospective and involved only the treatment group and self-report data. As mentioned above, there were very few comparative studies. The analyses in general were consistent with the methods employed. The samples were often convenience samples; meaning people who were easy to obtain data from and usually with a selection bias. Response rates were often low and non respondent studies were rare. Most of the instruments that were provided appeared to be sound but little information on the construction of the

instruments or indications of their validity was provided. There was little direct evaluator observation of programs.

All of the evaluations reported on how the program was operating and how that operation fit within NASA goals. They all also provided recommendations as to how the program might be improved or changed. The perceptions were overwhelmingly positive with a “glass half full” perspective. In terms of the summative questions there were very few evaluations that actually looked for pre to post change or comparisons. However, there were several retrospective questions that asked participants to comment on how much they felt they had changed. Again most people felt that the programs had affected them very positively.

It appeared from reading the reports that few if any of NASA decisions about educational programming was based on the evaluation reports. Three possible reasons for that come to mind. First, the evaluations may not have provided the information needed to make decisions. Second, the political environment moves much more rapidly than the evaluation environment and perhaps the reports were not available when decisions needed to be made. A third reason may be the very broad goals that are specified for the NASA programs. It would be very hard for any program, much less one with the limited funding available for NASA programs, to achieve these goals in any depth.

Introduction

This paper presents a summary of each of 10 evaluations conducted of NASA educational programs. The paper begins with a table outlining the titles of the evaluations and who conducted them, the date of the report, the evaluation questions, the evaluation design or methods and brief comments on the quality of each report. After the table each report is considered in more depth through an overview of what the evaluation included as well as a critique of the evaluation questions, methods and findings. The paper concludes with an overall commentary on the set of evaluations.

Table 1: Summary of NASA Evaluations

Evaluation	Date	Eval Questions	Design	Quality
<p>GLOBE Year 9 evaluation: Implementation Supports and Student Outcomes SRI</p>	<p>May 2005</p>	<p>How is GLOBE growing in terms of teachers trained and data reported? What do successful partners do that makes them successful? How do teachers find and use different GLOBE-developed materials, such as the Teacher’s Guide, when they implement the program? How do teachers experience and implement GLOBE? What is the effect of GLOBE on student achievement?</p>	<p>Existing data Case studies of two states Interviews with 7 teachers Matched comparison of two teachers’ students</p>	<p>OK and straightforward Good descriptions not nuanced Interesting data but limited generalizability Selection bias, teacher effect and limited generalizability</p>
<p>GLOBE Year 10 Evaluation SRI</p>	<p>September 2006</p>	<p>To what extent is the GLOBE Program growing in scale and reach, in terms of metrics traditionally used to measure growth? In classrooms where GLOBE is implemented, which aspects of implementation are important for improving students’ knowledge and inquiry skills? Can our assessment instruments reliably measure specific subscales of content and inquiry? What kinds of professional development activities in GLOBE are associated with increased levels of program implementation? What kinds of professional development activities in</p>	<p>There were three major studies. A quasi-experimental study of 45 classes for student achievement, a survey of teachers and providers about professional development and a case study of a field trial.</p>	<p>The studies were competently conducted although there were some design flaws mostly in the sampling. The issues presented are linked to the data. The recommendations extend somewhat beyond the data but appear justifiable.</p>

		<p>BLOBE are associated with increased teacher knowledge and changes to science teaching practices?</p> <p>How do support and follow-up after professional development influence program implementation and teacher knowledge and changes to science teaching practice?</p> <p>To what extent did GLOBE ONE achieve a balance of education and science goals, such that both educators and scientists could succeed in achieving their aims through the project?</p> <p>How well did GLOBE ONE's supports for student inquiry facilitate teachers and students engaging in their own investigations using GLOBE ONE data?</p>		
NASA Explorer Schools Evaluation Brief 1 Center for Educational Technologies	July 2003	Plan for a design experiment and a comparative study	No data	
NASA Explorer Schools Evaluation Brief 2 Center for Educational Technologies	February 2004	<p>What is the profile of schools designated as NASA Explorer schools</p> <p>What are the top target standards of selected schools</p> <p>What are the participants perspectives and beliefs about teaching learning and technology</p> <p>Who participated in the</p>	Surveys and perusal of applications and workshop agendas	<p>Straightforward presentation of the data, more base line. One needs assessment piece not really used.</p> <p>Some unnecessary information</p>

		<p>summer 2003 workshops and what did they do</p> <p>What was the participants feedback on summer workshops</p>		
<p>NASA Explorer Schools Evaluation Brief 3 Center for Educational Technologies</p>	<p>July 2004</p>	<p>What is the contextual background/conditions of participating schools</p> <p>How did the school teams organize to meet their goals</p> <p>How did school teams' strategic planning approaches work</p> <p>What is the quality of professional development supports</p> <p>How did overall NES program guidelines/supports facilitate participation</p> <p>What is the impact of program participation at end of year 1</p>	<p>Use of existing data, surveys of participants, focus groups of participants and program personnel</p>	<p>Good descriptions and formative feedback. A little wordy and tendency to make comments without clear supporting data.</p>
<p>NASA Explorer Schools Evaluation Brief 4 Center for Educational Technologies</p>	<p>February 2006</p>	<p>How is the NES Model Being Implemented?</p> <p>How does NES encourage more involvement with NASA program products and services?</p> <p>How does NASA involvement increase teacher professional growth?</p> <p>What is the effect of the program on school administrators?</p> <p>What is the effect of the program on family/caregiver involvement?</p> <p>What is the effect of the</p>	<p>Several different data gathering methods were used from three main perspectives: NASA personnel, Schools, and students and families. Methods included surveys, content assessments, interviews, observations, document analyses and interactions.</p>	<p>The report provided detailed information about the program. The effect data was mostly perceptual and there was selection bias. Some statistics seemed inappropriate. More precision in reporting the results and more linking of evidence to evaluative statements would have improved the report.</p>

		<p>program on students' interest, career aspirations, and knowledge of science, technology, engineering mathematics and geography?</p>		
<p>NASA Explorer Schools Evaluation Plan 2006-7 Paragon TEC</p>	<p>November 2006</p>	<p>Plan for an evaluation</p> <p>Overall Question</p> <p>What is the relationship of the nature and extent of a school's involvement to their success in developing teachers' competence in using NASA STEM-G resources and student interest, attitude and achievement in STEM-G</p> <p>What is the nature of an NES school's use of NASA resources</p> <p>What is the extent of an NES school's use of NASA resources</p> <p>In what ways and to what extent do the short duration professional development activities associated with being a NASA Explorer School affect teachers' confidence, competence and use of NASA for STEM-G instruction</p> <p>In what ways and to what extent do the long duration professional development activities associated with being a NASA Explorer School</p>	<p>No Data, Proposed NEEIS, other surveys of teachers and students, student content tests for selected students, surveys of staff</p>	<p>No Data Proposed intensive comprehensive data collection effort, data focus on perceived and actual effect on participants, some pre post data, no strong comparative information</p>

		<p>affect teachers' confidence, competence and use of NASA for STEM-G instruction</p> <p>In what ways does NES involvement affect family involvement In what extend does NES involvement affect family involvement</p> <p>To what extent does NES involvement affect student interest in STEM-G topics To what extent does NES involvement affect student attitude toward STEM-G careers</p>		
<p>Sun-Earth Day 2006—Eclipse: In a different light Technology for Learning Consortium Inc</p>	<p>September 2006</p>	<p>Evaluation was to examine the following goals Inspire large audiences with NASA/Sun-earth connection science and current events Support the understanding of fundamental Sun-earth connection themes Facilitate participation by education and public outreach coordinators in sun-earth day</p>	<p>Surveys of participants, tracking of use of web sites and materials</p>	<p>Straight forward perceptual and use data</p>
<p>NASA International Space Station EarthKAM Program EDC</p>	<p>November 2006</p>	<p>To evaluate the program against the NASA educational goals and provide strategic recommendations for future directions</p>	<p>Interviews of project staff and participants; use of NEEIS data; site visit to UCSD</p>	<p>The data are limited but there are some directions. The conclusions aren't as related to the data as they could be</p>
<p>Digital Learning Network Evaluation Tool Development Technology for</p>	<p>November 2006</p>	<p>Develop an assessment device for the reduced gravity module Develop a rubric for assessing the quality of DLN modules</p>	<p>Content assessment test</p>	<p>The test was in the process of being improved. No comparative data were provided. Items appeared to be mostly fact based</p>

Learning Consortium Inc		with extended definitions		
Evaluation of the NASA Aerospace Education Services Program The Evaluation Center	October 2004	There are 19 evaluation questions addressing the following 5 areas Program design and management, Support of systemic improvement, Teacher preparation and enhancement programs that support systemic reform, Student support and Curriculum and dissemination	Delphi survey, Surveys of an AESs and CPOs Telephone Interview Protocol, AESP State Impact Survey, AESP and CPO NASA Explorer School Surveys interviews and site visits document review, NEEIS data,	The report is carefully done and presents a great deal of data in an easily understandable fashion. The data are more descriptive and perceptual.
The Final Report of a Study of the Aerospace Education Service Programs (AESP) Role and Impact Among Selected Partners The Evaluation Center	April 2006	With whom does AESP cooperate and support for delivery of NASA programs to students, teachers and others? What is the form and nature of this cooperation and delivery of services? How effective is AESP in its provision of support services for its NASA and non-NASA partners? How do these cooperative actions and provision of services to other NASA partners impact on the traditional role of AESP What are the elements or activities of AESP that contribute most to NASA's major education goals? What are some exemplary cases in which AESP specialists' work has impact?	Site case studies, surveys, NEEIS data	All of the data are limited by small numbers of respondents. The case studies could have focused more on NASA value added. The conclusions are not always directly supported by the presented evidence.

GLOBE year 9 Evaluation; SRI; May 2005

This extensive report provides information on four different substudies, each designed to address a different evaluation question about the GLOBE program.

- How is GLOBE growing in terms of teachers trained and data reported?
- What do successful partners do that makes them successful?
- How do teachers find and use different GLOBE-developed materials, such as the Teacher's Guide, when they implement the program?
- How do teachers experience and implement GLOBE?
- What is the effect of GLOBE on student achievement?

How is GLOBE growing in terms of teachers trained and data reported?

This question was addressed using existing data bases about number of teachers trained, numbers reporting data and numbers reporting data over time. These data showed that although the overall numbers of teachers trained was relatively stable, this was due to a decrease in teachers trained internationally and an increase in teachers trained in the US, most predominately in Alabama. The reporting data showed fluctuations over time commensurate with the school year (reporting down in the summer) and that only a small number of the teachers trained actually reported usable data (considered honor roll teachers). Reporting also varied by topic with atmosphere and hydrology highest, distantly followed by Soils. Teachers persistently reporting data for one year and two years was also reported. This showed that some teachers were consistently engaged. The trends appear to show that there are some very committed teachers and others who are less so.

The presentations of the data were straightforward and well done. The authors are careful to point out limitations and contrary trends in the data. It would have perhaps been better to have used percentages in some places rather than direct numbers. The authors suggest using percentages themselves but don't explain why they didn't do so in this report. They also don't explain why they only present persistence data for only two years. This seems a limited amount of time given this is a year 9 report. The conclusions are based directly on the data but do present things in a "glass half full" tone. Tracking these sorts of data seem like an important and appropriate evaluation task.

What do successful partners do that makes them successful?

This question was addressed through case studies of two of GLOBE's successful US partners. The authors present these case studies in a format that provides hypothesis and alternative hypotheses. They suggest that they used the one case to refine the hypotheses and then re-examined them in the second case. The report provides a fairly detailed description of what the partners do. It also considers the areas of the hypotheses and issues related to them. The conclusions are that the two partners operate in different ways but do provide extensive support to the teachers. Also pointed out is the fact that the teachers themselves provide considerable support to facilitate continuation of the program.

Although the report provides a good description of the two sites that illuminates the commonalities and differences, a reader is left wishing there were more nuanced analyses. Additionally there are many potential rival hypotheses that were not addressed and actual changes to the hypotheses were not highlighted. It seemed like the report began that way but then settled into a more typical case study format. There is little emphasis on how the partners actually obtain the funding necessary to support their operations. A more pointed analyses of how the partners managed to integrate themselves with ongoing State projects and the amount that contributed to their success would have been helpful. As it is, the report offers little advice for other partners. The notion of teachers providing support is important and the descriptions do highlight teacher efforts but the report doesn't really indicate what might have made these teachers be supportive in the ways they were so that this could inform the GLOBE program how to help other teachers become as supportive as these are.

How do teachers find and use different GLOBE-developed materials, such as the Teacher's Guide, when they implement the program?

How do teachers experience and implement GLOBE?

These questions were addressed with a small study of teacher use. The study used only six teachers from northern CA and one from Queens, NY for a total of 7. All teachers were heavy users of GLOBE. There was a spread of grade levels. Teachers completed an on line survey asking about their GLOBE use and were then interviewed at their school about their use of one or two of the protocols and activities they reported using. These were the Cloud and Temperature protocols for the elementary school teachers. The two high school teachers discussed Hydrology and Soil. When the teachers were describing their used of the protocols they commented that the program was better than textbooks and the opportunity to really touch soil etc was worth while. The conclusions were that the teachers used the materials differently depending on their purposes, that they thought the data were useful to scientists and that they integrated GLOBE with other activities. There was a long list of suggestions for improving the materials.

This small study seemed useful although it might have been just as effective to conduct a few focus groups at sessions with a variety of GLOBE teachers in attendance. This sample was very small and localized. The inclusion of one teacher from NY seemed somewhat out of place. The inclusion of different grade levels was appropriate but it might have been better to link the selection to the percentages of teachers of these types in the project. The authors make comments that suggest more generalizability from the sample than may be justifiable. The authors also suggested that "A greater understanding of the merits of different approaches to introducing GLOBE could be a future area of study." It seems that a study of this type might have been conducted sooner than nine years into the program. Many of the teachers' comments appeared to relate to hands on learning as opposed to specific elements of GLOBE except perhaps a general connection to scientists. There seemed to be quite a bit a difference in the responses of the elementary and high school teachers that were not separated out too well.

What is the effect of GLOBE on student achievement?

This question was addressed with a matched comparison study of two teachers and four classes of their students during a unit on Hydrology. There were 123 students they were tested on their understanding of hydrology concepts and their attitudes. Teacher perceptions were also assessed.

The study had several design limitations. The effect due to teacher was confounded with the effect due to the curriculum. Although the teachers were presented as similar there could easily have been differences between them. Most notably one teacher had been interested enough to participate in GLOBE, so there was a selection bias. There were no data that really documented how different the classes really were, although the authors say they were comparable. The GLOBE students had higher achievement scores to begin with so there may have been differences between the classes even though they appeared to be matched. It was not exactly clear how the pretest and posttest were used. The authors mention gain scores as well as ANCOVA so it appears that they subtracted the pretest score from the post test score to obtain a gain score and then also used the pretest score as a covariate. They find that the pretest scores are significant so the fact that the GLOBE students had higher pretest scores might be even more of a problem in terms of measuring effects. The attitude scales were formed without factor analyses but they did have reasonable reliabilities. Apparently there was no pretest on attitudes and these were post test only comparisons between groups. It is not clear why a pretest was not given. The authors say that this was the strongest findings to date.

Conclusion

These individual studies were designed to answer the evaluation questions. The questions themselves appear relevant but the data and the analyses presented are not as comprehensive as they might have been. The existing data were used well and the charts were informative. The case studies were thorough and involved the collection of considerable amounts of data but the actual analyses were not as directed to answering the evaluation question as they might have been. The interviews with the seven teachers provided reasonable formative information but the sample was very restrictive and therefore the data were less valuable. The final matched comparison study had several flaws so the information was not particularly useful. The instruments used seemed of high quality and the authors used the techniques appropriately. They reported carefully and pointed out most of the limitations in their data.

The recommendations that the authors provide show more insight than is presented in the individual studies suggesting that more data than was reported on was probably available.

Overall for a report that is supposed be of the 9th year of a project, the data and analyses seemed somewhat superficial.

GLOBE Year 10 Evaluation; SRI; September 2006

This is the report from the 10th year of the GLOBE program the 2004-05 year. There were several evaluation questions including:

- To what extent is the GLOBE Program growing in scale and reach, in terms of metrics traditionally used to measure growth?
- In classrooms where GLOBE is implemented, which aspects of implementation are important for improving students' knowledge and inquiry skills?
- Can our assessment instruments reliably measure specific subscales of content and inquiry?
- What kinds of professional development activities in GLOBE are associated with increased levels of program implementation?
- What kinds of professional development activities in BLOBE are associated with increased teacher knowledge and changes to science teaching practices?
- How do support and follow-up after professional development influence program implementation and teacher knowledge and changes to science teaching practice?
- To what extent did GLOBE ONE achieve a balance of education and science goals, such that both educators and scientists could succeed in achieving their aims through the project?
- How well did GLOBE ONE's supports for student inquiry facilitate teachers and students engaging in their own investigations using GLOBE ONE data?

The answers to the questions were grouped into three main studies. Questions 2 and 3 were addressed using a quasi-experimental study in 46 classrooms. Questions 4-6 about professional development were answered using HLM analyses of teachers nested in programs with different outcome variables. This included 454 teachers from 28 of the most active GLOBE partners. The data were surveys of teachers and information from program providers. The final questions were answered with a case study of the GLOBE ONE field campaign.

Overall this GLOBE report was much more precise and carefully conducted than the GLOBE 9 report. The data were presented in more detail and the analyses were more sophisticated. There was also connection to existing research. In general the authors are careful to present their data clearly and to point out limitations. The format also makes the report very easy to follow.

The tracking data are much the same as in GLOBE 9. The data show similar trends and are only briefly presented here, justifiably so since it is generally a marginally useful indicator of program success. One interesting item of data that is not pointed out is the increased use of automatic data inputting systems. This somewhat artificially increases the records of reporting since it is not the students and teachers doing it.

The experimental study had several flaws, most of which the authors pointed out. There was really no control on the teachers. They selected teachers who were using the GLOBE materials and GLOBE trained teachers who were not supposed to be using the materials. The authors point out that the use of materials did not necessarily fit the predetermined categories but they somewhat gloss over the fact that the teachers expected to not use GLOBE could be systematically different than the others. They, in fact, suggest that the teachers were the same in explaining the lack of significance in their results. More controls on the teachers or random assignment is necessary. Also the

authors suggest that their instruments may not be as consistent as ideally hoped for. The instrument itself is somewhat eclectic and details of the construction are not provided—just the lack of internal consistency. Use of existing tests or items might have been more productive. If NAEP or TIMSS items had been used, then percent correct could have at least been compared to national averages.

The professional development study was based on another comprehensive look at professional development in the literature and so added significantly to the field. The authors are careful to point out that it is a correlational not a causal study but the results are indicative of theoretically sound relationships and provide good indications of what might be effective. The sampling was less than ideal. They said they sampled more people than they needed because they expected a low response rate. This is not really acceptable procedure because it really assumes a bias in the respondents. The authors did conduct a non-respondent study but it was only on school demographic variables and there is likely to be more variation among teachers than among schools.

The case study was conducted using standard procedures and the reporting of the findings appeared to be unbiased. There was a good grounding of the study in research and negative results were presented. They didn't really use the grounded theory and hypothesis checking they suggested in the introductory section but then grounded theory and hypothesis checking are necessary in case studies.

The conclusions presented are clear and directly linked to the data although the student achievement conclusion is a little more positive than in the actual report. The division between the recommendations and the conclusions clearly shows that the recommendations go beyond the data which is indeed the case.

NASA Explorer Schools Evaluation Briefs 1, 2 and 3; Center for Educational Technologies; July 2003, February 2004 and July 2004

These three documents form a set.

NASA Explorer Schools Evaluation Brief 1; July 2003

The first brief is a description of the proposed evaluation plan for the Explorer schools. The plan is to use the evaluation as part of a “design experiment” to improve the program. Design experiments are generally a series of attempts to address a problem where different hypotheses of how the problem might be solved are tested and modified. Eventually an answer to the problem is determined. In addition to the formative design experiment approach, the evaluation brief proposes an experimental design for a summative evaluation where Explorer schools are compared to other schools. The Brief outlines the NASA goals for the program and the three year plan for its development.

NASA Explorer Schools Evaluation Brief 2; February 2004

The second Brief is a report on the selection of the Explorer schools and on the summer workshop. This report includes a summary of much of the material presented in Brief 1

in addition to several different types of data. The authors suggest five evaluation questions:

- What is the profile of schools designated as NASA Explorer schools
- What are the top target standards of selected schools
- What are the participants perspectives and beliefs about teaching learning and technology
- Who participated in the summer 2003 workshops and what did they do
- What was the participants feedback on summer workshops

The authors use data from review of application materials and workshop agendas and from administration of several different types of surveys to answer the questions. The surveys include both author developed ones and ones available in the literature. The authors also use the NEEIS data on NASA participants. They use a framework from the literature on professional development to organize the information about the activities.

The review of the application materials is very straightforward and provides information on the type of schools and the team participants. The author developed survey of what standards the school teams target showed that there was some agreement across the teams in terms of goals. This survey was designed to be used as a needs assessment but it does not appear to have been used that way.

The teachers completed five scales of the Teaching Learning and Computing Teacher's (TLC) Survey. This will provide base line information about the teachers skills, attitudes and constructivist teaching. The authors conducted t tests to determine differences from the middle score of 3 which seemed unnecessary. The mean scores themselves were sufficient.

The sample was described using self report data and the different components of the workshops were presented according to structural and core features. This distinction was literature based and was useful but not completely necessary. The authors also considered all of the professional development that might be provided not only the summer workshop and compared this to "typical" professional development. Again this was useful but not completely necessary and at times somewhat confusing. For example they presented several tables of the percent time spent on various things but there was not an indication of the total time spent on the components across tables. There was also some inconsistency between what labels were used in the tables and which were used in the narrative.

The perspectives on the summer workshops were provided from the NEEIS data which included a variety of formats for determining satisfaction. This made for a somewhat confusing presentation of different types of data but a picture of satisfaction emerged.

The report ends by suggesting what might be the implications for seeking and sustaining coherence. The view of the evaluation is that the school plans and the professional development should be coherent in terms of design and implementation. The authors

point out that it is too early to actually determine coherence but they make several suggestions as to how to move forward.

The data presented in the report are clear but the tone is a little more positive than the data warrant. For example, they suggest findings indicate the potential for success when that cannot really be determined. The authors are very clear, however, about the tentative nature of their findings since this is only the beginning of the program.

NASA Explorer Schools Evaluation Brief 3; July 2004

This report presents the results for the first year of the program. It also includes summaries of the results from Briefs 1 and 2. Data in the report is based on information from the program about meetings and attendance, etc from the NEEIS and on the results of several focus groups held at various events with participants and field center staff. The teachers also completed the TLC survey again and the school completed a survey of technological capabilities. The report also presents a logic model for the program and subsequently for the evaluation. There were six evaluation questions:

- What is the contextual background/conditions of participating schools
- How did the school teams organize to meet their goals
- How did school teams' strategic planning approaches work
- What is the quality of professional development supports
- How did overall NES program guidelines/supports facilitate participation
- What is the impact of program participation at end of year 1

The findings include findings from the summer but are supplemented by findings during the school year that shed light on the longer term effects of the summer workshop. The perspective is somewhat less positive than in Brief 2. The data from the focus groups are organized to answer the evaluation questions. The narrative shows good insights and the quotes are illuminating but the actual extent of some of the issues raised is difficult to gauge since numbers are not provided. The technology survey shows that the schools are not particularly well equipped. The TLC survey showed positive gains for some of the scales. The descriptions of the special events and the attendance record is useful and displayed well. The final insight, though, that schools that were well organized were most likely to report positive results seems a little superficial but perhaps all that could be said given that the data were only from focus groups. The conclusion section provides 8 lessons some linked to existing research.

It appears from this report that the design experiment idea is not working. There doesn't appear to be any concerted effort to design or even monitor differences in planning or implementation in a way that would allow development and testing of hypotheses. They do report, however, some changes in the program in relation to identified weaknesses. The report is useful in that it indicates several issues that will need to be addressed in order for the program to be successful.

NASA Explorer Schools Brief 4: Evidence that the model is working; Center for Educational Technologies; February 2006

This was a very comprehensive evaluation report containing information from several other reports and building on the prior Briefs. This report continues some of the data collection begun earlier but also presents new information. The evaluation questions are:

- How is the NES Model Being Implemented?
- How does NES encourage more involvement with NASA program products and services?
- How does NASA involvement increase teacher professional growth?
- What is the effect of the program on school administrators?
- What is the effect of the program on family/caregiver involvement?
- What is the effect of the program on students' interest, career aspirations, and knowledge of science, technology, engineering mathematics and geography?

Several different data gathering methods were used from three main perspectives: NASA personnel, Schools, and students and families. Methods included surveys, content assessments, interviews, observations, document analyses and interactions. NASA personnel data included: NES coordinator weekly activity reports, coordinator interviews, AESP NEEIS data, field center surveys, workshop agenda analyses, observations, and center case studies. School data included: team focus groups, Teacher NEEIS data, teacher content knowledge growth (this was mentioned in the report but the data were not yet available), number of participants, teacher attitude and skill growth, teacher technology TLC survey, Teacher action research, school case studies (2 per center). Student and family data included: student symposium results, student content knowledge (this is mentioned in the report but the data were not yet available), number of participants, career survey, family survey (this was mentioned in the report but the data were not yet available), school case studies, student products.

The results provide a detailed description of what is going on in the program and the participants' perceptions of the activities. A summary of the results from the prior Briefs is also included in the introduction. The data collection activities had some limitations. The Teacher focus group interviews were conducted with only the 2003 cohort teams and were conducted via phone rather than in person and in 10 of the 49 cases were conducted with only the team lead. The numbers of NASA personnel responding to requests for their perceptions does not appear to be in the report.

The Observation of the Workshops scheme was based on prior research and seemed appropriate but the results were difficult to translate into formative or summative information about the workshops. They were more status oriented. Also the tables presenting these data were very difficult to understand. It appeared that they might have been reporting the percent of materials/activities that were used in the workshop that had had no emphasis, minor emphasis or major emphasis on the different topics. However, the narrative did not consistently present the data in that way so that interpretation may not be correct.

The teacher surveys were given to teachers who attended the workshops and the TLC was obtained from 41 to 77% of the teachers. This might indicate some bias in the respondents in favor of more involved teachers. The administrator survey was given to only the administrators who were on the teams so there is a built in bias. Of the 50 schools per year 44 42 and 34 administrators returned surveys. The Student Symposium Student Survey also had a built in bias in that these were the students chosen to attend this event and so were not necessarily representative of the students in their schools. The Student Interest Survey was given to students in general.

The report provided a great deal of data and apparently there were even more detailed data in the Appendices but those were not available to review. The executive summary provided an overview of the conclusions derived from the study and the statements appear reasonable and useful. However in the summary there is no indication of what evidence the conclusions were based on. It is difficult to provide all of the evidence in a summary but something about the sources from which the conclusions were derived would have been useful. There is also no discussion of the possible limitations of the evaluation.

Although the report is organized to focus on two major questions and the data are presented by detailed evaluation question, the report is still quite dense and it is somewhat difficult to focus on what is really the evidence. This is exacerbated by the fact that almost all of the questions are answered using data from several sources. Although this sort of triangulation is good in providing a well rounded picture of the program, the authors are not always careful to specify which data they are using to support their statements. For example, "teachers said .." could have been data from many different sources each with different bias. Or statements like "it was found.." were very vague with no indication of where the evidence came from. The authors were also not very precise in their statements which led to over generalizing the results.

The descriptive results are interesting and provide a picture of what is going on in the schools and the NASA centers. It seems though that much of the information about programming would be already known although perhaps it is important to have this documented externally. The results about operation really match the school improvement literature. It would have been nice to have more detail on the NASA parts. What is different about a NASA school improvement than just a general one? There were hints at this in the NASA prestige and need for AESs to plan events but many of the other things were to be expected, e.g., increased teacher competence is related to an organized team lead that has the time to communicate and organize the connection. The analysis of the workshops suffers from the problem noted above with the data collection device and the presentation method. The analysis of the school plans presents detailed information about what the schools planned and could be valuable if it were compared with some of the other data but it is just presented not really compared to other sources. The authors did apparently do some unreported comparisons because the recommendations relate to other aspects of the program.

In the analyses of the effects on teachers, administrators, families and students the data seem to be analyzed in ways that might not be as accurate as should be. The authors report on teacher changes but they report data that are not necessarily matched. They look at changes in cohort 3 in two different years but they don't match the teachers so the changes could be due to the different samples. They do sometimes provide both a paired and a non paired analysis so that is useful. They also conduct what they call is an analysis of covariance between various teacher scores but it seems to be more just correlations. Also there seem to be mistakes in the tables (missing marks for significance levels and unclear notes). The administrator surveys come from three different cohorts but the authors suggest that differences in the cohort scores might be related to maturation effects. This seems reasonable but the differences could just as easily be due to the different samples. To really show a maturation effect, they would have had to measure change on the same people. These tables are also not very clearly labeled as to what the numbers in the tables represent. The family data is all perceptions of people who are trying to engage family members so it may be suspect. The effects on students suffers the most from not being clear as to what data source the comments are related to. Data from the students at the symposium are much less representative than the student interest survey and it is not always clear which data are being cited. Again the authors don't match the students but still talk about change using year cohorts. Some of the career data appear to be more linked to the symposium students so the positive results may be somewhat related to the selection bias of which students attended the symposium. Other career data are on the student interest survey but it appears that the students where asked to select from a list of science careers not from a list of any career so those percentages are somewhat biased as well.

NASA Explorer Schools Evaluation Plan 2006-7; Paragon TEC; November 2006

This is a modified evaluation plan for the next phase of the Explorer Schools. The authors describe the prior evaluation work in a summary and then detail what data the new evaluation will collect. The evaluation questions are listed below.

Overall Question

What is the relationship of the nature and extent of a school's involvement to their success in developing teachers' competence in using NASA STEM-G resources and student interest, attitude and achievement in STEM-G

- What is the nature of an NES school's use of NASA resources
- What is the extent of an NES school's use of NASA resources

- In what ways and to what extent do the short duration professional development activities associated with being a NASA Explorer School affect teachers' confidence, competence and use of NASA for STEM-G instruction

- In what ways and to what extent do the long duration professional development activities associated with being a NASA Explorer School affect teachers' confidence, competence and use of NASA for STEM-G instruction

- In what ways does NES involvement affect family involvement
- In what extend does NES involvement affect family involvement
- To what extend does NES involvement affect student interest in STEM-G topics
- To what extent does NES involvement affect student attitude toward STEM-G careers
- To what extend does NES involvement affect student achievement in STEM_G

Data are to be collected through extensive use of the NEEIS data system, eportfolio, student and teacher surveys and surveys of staff. There also appears to be a subset of schools designated as case study schools but the only additional information appears to be a request to the schools to try and collect student achievement data using a randomized control trial.

The plan is very detailed and includes tables linking questions and instruments to NASA goals. The plan seems to provide very detailed information about what everyone has been doing and their perception of the activities. Despite the detail, however, it is not clear how the overall question about school planning and its effect on teacher professional development and student outcomes will be answered. How the links will be studied is not explained in detail. It seems to be a very labor intensive effort on the part of the local people in filling out forms and in administering and collecting surveys. The data may not allow the comparison of short and long term professional development suggested in the report because it appears to me that participation in long and short term opportunities is confounded. Also student data are restricted to selected sites with only suggestions of how to actually design the local experiments. There is also no mention of hierarchical analyses where students would be nested within classes and then within schools to allow the examination of the effects of these different levels. This would be possible with the 25 schools specified and would allow some indication of the effect of variables at the different levels. Additionally there is no attempt to compare the results of these schools with other schools in the state or nation to provide information other than that if you teach it things change. Use of the state tests or NAEP data in some way might improve the opportunities for comparisons.

Sun-Earth Day 2006—Eclipse: In a different light; Technology for Learning Consortium, Inc.; September 2006

This was a brief evaluation report combining the data from several different surveys and tracking mechanisms into answers about whether or not the program met the following program goals.

- Inspire large audiences with NASA/Sun-earth Connection science and current events
- Support the understanding of fundamental Sun-earth Connection themes

- Facilitate participation by education and public outreach coordinators in sun-earth day

Data collection included the administration of surveys to different types of participants (Museums, classroom educators, NASA Mission Education and Public Outreach leads, scientists and amateur astronomers at their various venues. Requests for materials and web connections and downloads were also tracked. Participants were also asked about ways the program could be improved.

The results show very wide participation and high perception of engagement and knowledge growth and that the coordinators had high levels of participation.

This evaluation provides a very straight forward and accurate look at perceptual and use data. The authors are careful to present the evidence that their claims are based on. The only flaw is the selection bias. All people included were people who wanted to participate (except perhaps the coordinators) so there is no information as to why people may not have participated or how to motivate people not interested in these types of programs to participate.

NASA International Space Station EarthKAM Program; Education Development Corporation; November, 2006

The goal of this evaluation was to consider the EarthKAM program against the NASA educational goals and to provide strategic recommendations for future directions. Data were collected from face to face telephone interviews with five teachers (one of which was an add on), observations and field notes from a site visit to UCSD, interviews with NASA staff and students at UCSD and data from the NEEIS system. The report provides a description of the program's components and operations and a discussion of the program's effectiveness in relation to NASA goals. The data from the NEEIS system is generally presented clearly and shows involvement over time although it wasn't always clear what the data meant. The report highlights the value of the program to the UCSD students.

The authors were very clear about the limitations of their evaluation. They only had four teachers to interview and the data from the NEEIS system were not readily available in formats that would allow data analyses. Despite pointing this out they still present data from the four teachers in ways that indicate that they are representative of the program overall. The report presents the data and makes some suggestions about its effectiveness. The conclusions are somewhat weak perhaps mostly because the results were somewhat negative and the authors didn't want to point that out explicitly. There are some sections in the conclusions that don't seem to be directly related to the data.

Digital Learning Network Evaluation Tool Development; Technology for Learning Consortium, Inc.; November 2006

This report is the description of the development of an assessment instrument for the Reduced Gravity Module. It also presents a detailed rubric for assessing the quality of DLN modules across 14 criteria with scores of 0-4. The rubric provides descriptions of what each rating would mean for each of the criteria. In addition the report provides detailed explanations of each criterion so that developers would understand the criteria and what would be required to score at each level. Data from the assessment are presented for 69 K-6 teachers and 78 sixth grade students both before and after engaging in the DLN module. Both groups showed significant improvement by item. The data were used to conduct item analyses and suggest improvements for the assessment instrument.

The rubric and its detailed explanations are reasonable. It provides a good deal of science education information in a small space that could be helpful to DLN investigators. There is no data on the usefulness perceived or otherwise of the rubric. The assessment results show that people knew only a little about the effects of reduced gravity before participating in the module but much more afterwards. This is good news but doesn't address the comparative advantage of the module. Also the instrument itself still need refinement. The authors point out different foils that need to be improved. Additionally reliability analyses could be conducted. Finally the items seem to be mostly of the fact variety rather than application or analyses even though the rubric explanations are quite clear about three different levels of cognition that questions should address.

Evaluation of the NASA Aerospace Education Services Program; The Evaluation Center; October 2004

This was a comprehensive, three year evaluation of the Aerospace Educations Services Program (AESP). This final report also incorporated results from two earlier reports; one on the Delphi study and one on the survey of AESs and CPOs. There were 19 evaluation questions which guided the report. These were grouped into five areas.

Program Design and Management

- What problems or needs is the NASA-AESP attempting to address?
- How effectively is the program designed to meet the needs?
- To what extent are the participants/program activities meeting the anticipated needs of the program?
- What strategies/activities are most effective in aiding the participants to meet the program needs, and what barriers are being encountered?
- How effectively are resources (human and financial) being deployed to meet the needs of the program?
- What significant unintended impacts/outcomes did the program have?

Support of Systemic Improvement

- How and to what extent, did the program and activities comply with state and local guidelines and frameworks and contribute to support of state-level systemic education improvement efforts?

- How effectively were programs and materials that included standards and curriculum frameworks as core elements developed and used?
- How have program activities impacted targeted K-12 students?
- How and to what extent, did the program activities impact on classroom teachers and their instructional planning and practice?
- What benefits are derived from close relationships with NCTM, ITEA, NSTA and NCGE; and what is the importance of these benefits in fulfilling the goals of the overall program?

Teacher Preparation and Enhancement Programs that Support Systemic Reform

- To what extent and how effectively were teacher workshops developed and offered that supported state-level issues and initiatives?
- To what extent were teacher workshops designed to include NASA curriculum materials, state-of-the-art educational technology, and pedagogical skills that integrated the NASA mission as a common element with the subject matter content?
- How did these workshops impact teacher practice?

Student Support

- To what extent did the program increase support to preservice teacher education programs at colleges and universities?
- How effective was the program in increasing the number and value of workshops for preservice educators

Curriculum and Dissemination

- In what form, and how effective were program efforts to support informal education in institutions and organizations, teacher training programs, and other providers of related services?
- To what extent did the program stress the inclusion of aerospace information in support of state-based initiatives?
- How effective was the program in helping the Urban and Rural Community Enrichment Program focus on the interests of minority teachers, students, and members of the general public in the targeted areas?

A variety of methods were used to gather data from a provider and a client group. All of the providers were asked to respond to data collection requests and a sample of clients. Data collection devices include a three round Delphi survey, an AES Effectiveness and Working Conditions Survey, a CPO Perceptions of AES Effectiveness and Working Conditions Survey, Telephone Interview Protocol, AESP State Impact Survey, AESP NASA Explorer School Survey and the CPO NASA Explorer School Involvement survey. Other data collection activities included interviews and site visits to NASA field centers and to local schools where AESP specialist made primary presentations to students and teachers, document review, examination of the NEEIS data, Observations of AESP professional development activities and AESP specialist presentations. The overall

conclusions are that the program is generally effective although more and less so in different areas and that the AESP are quite dedicated and productive.

The report presents all of the data collected to answer each of the 19 evaluation questions. The report is very effective highlighting important points, pointing out limitations of the data and clearly reporting negative results when they were warranted. The distinction between provider and client was good for differentiating the source and lenses of the data and also allowed more insightful interpretations of the data. Unfortunately the distinction wasn't carried out in as much depth as could have been done throughout the report. There are some client provider comparisons which are good but sometimes a little difficult to follow and statistics aren't used consistently. The report provides a nice historical contextualization to ground the study. Throughout the report data from different sections are cross referenced which provides a much more comprehensive picture of the program. The idea of calling rather paper and pencil surveys for the clients was a sound idea to obtain better response rates although even with that the number of clients responding was quite low. The inclusion of evaluation questions about program management was a good idea to provide insight into program operation. The time spent/efficiency analyses were quite informative. There were a wide variety of data collection procedures which provided a comprehensive look at the program but the data were largely descriptive or perceived impact. The instruments themselves were of high quality although a little dull. The authors provide good evaluative interpretations of the data but in a few instances the data supporting the insights are not directly provided. These interpretations probably arise from synthesis of all of the data given that the evaluators had been immersed in the program for such and extended time. The authors provide somewhat of a look into the future of the program but the conclusion section is weak.

The Final Report of a Study of the Aerospace Education Service Programs (AESP) Role and Impact Among Selected Partners; The Evaluation Center; April 2006

This evaluation was a supplement to the report in 2004. There were six evaluation questions.

- With whom does AESP cooperate and support for delivery of NASA programs to students, teachers and others?
- What is the form and nature of this cooperation and delivery of services?
- How effective is AESP in its provision of support services for its NASA and non-NASA partners?
- How do these cooperative actions and provision of services to other NASA partners impact on the traditional role of AESP
- What are the elements or activities of AESP that contribute most to NASA's major education goals?
- What are some exemplary cases in which AESP specialists' work has impact?

This report includes the results of four studies. The first provides detailed descriptions of eight sites of the 10 total that had been nominated by AES's as exemplary. The second

was a survey of the AESPs to identify on and off center partners with whom that had worked and what the nature of these interactions had been. There were 21 respondents from the 38 surveyed. The third study was a follow-up with the partners identified in study two. The fourth study was analyses of information from the NEEIS system about the activities of five AESs in three past years.

Each study is well conducted and the report provides good information but it is less well organized than the 2004 report. The report takes a “lets look at the best” approach for the studies and all are limited by the small number of respondents. The bulk of the report is the site reports which provide nice examples of what is going on at the sites. The descriptions present a lot of “opinion” about the impact and discuss many things that aren’t directly AES related to provide the necessary context. They might have been more directly focused on NASA value added. The results of the studies seem to highlight unique individual commitment and local context which isn’t highlighted as well in the conclusions. The quotes and data presentations are generally good but the ones on percent time are too detailed and don’t present the full picture. Some of the summary conclusion statements are less well supported by the evidence provided than others. The report also addresses to some extent the effect of the new Explorer School approach on the AESs.

Conclusions

Section 5 Planning and Evaluation discusses both formative and summative evaluation and provides several suggestions of what may be involved in each. It is suggested that formative type questions include those such as the following: Are the activities that the teacher and students are engaged in consistent with those intended in the program? Is the content that engages the teacher and students consistent with the intended content? Do the teacher-student interactions and the student-student interactions match those intended? Is the nature of the classroom discourse consistent with that envisaged in the program? The Planning and Evaluation section also list several types of potential data gathering devices (e.g., surveys or teacher logs) and methods (e.g., classroom visits) as well as data analyses and interpretation possibilities. The section recommends that both qualitative and quantitative data be analyzed in a scientifically justifiable manner and that instruments used to collect data should be first examined for their reliability and validity. Further it recommends that the results of the data analyses should be displayed clearly in tables or graphs with discursive summaries.

The Planning and Evaluation section views summative evaluations as slightly different from formative ones but states that they follow the same steps. Although summative evaluation of any education program can be designed to address different questions, there are two major ones. One is to determine whether, and to what extent, the program results in the desired outcomes as specified in the program. The second question asks whether teacher and/or student outcomes are the same or different (better, worse) when the NASA program is compared to other science education programs. In other words, it is important to find out what changes are occurring in teachers’ practices and attitudes and subsequently in their students. It may also be important to perform comparison studies but these often have difficult design requirements. Data gathering devices can be similar

to those suggested for formative evaluation but also include achievement tests and surveys of other outcomes. Because of widespread interest in the results of summative studies, a variety of dissemination techniques is recommended.

Therefore the question for this critique becomes how do the NASA evaluations of their educational initiatives fit within the guidelines suggested in the Section on Planning and Evaluation. It is difficult to classify the reports as summative or formative. They all appeared to have components of each. For example, the NASA International Space Station EarthKAM Program evaluation goals are to evaluate the program against the NASA educational goals and provide strategic recommendations for future directions. Therefore the critique will focus on methodology (design, data collection methods, analysis and reporting.)

The reports were all more or less typical evaluation reports seemingly intended for only internal use. There was no evidence of any of the reports being disseminated to different audiences or with different techniques. The reports themselves were generally of acceptable quality with appendices presenting detailed information and brief, but informative, extracts of the data provided in tables in the body of the report. There were only a few instances when the tables were difficult to interpret. The narratives surrounding the data were usually clear but often imprecise. The narrative would often include unproven inferences from the actual data or gloss over subtle differences. This included not always being clear about whom the data were actually from or the numbers of respondents a particular bit of data referred to. In the longer reports the data from the various sections were often not cross referenced so the reader had to go back through the report to do comparisons herself. This was often true for the information related to the sample. Many reports had a section at the beginning that detailed the various techniques and samples but this information was not always repeated in the section where the data were presented.

The analyses in general were appropriate to the methods employed. Quantitative data were analyzed statistically while qualitative data were analyzed interpretively. Some exceptions were that the assumptions of the statistical techniques were not always met. In a few cases samples were too small for the analyses conducted or the statistical technique seemed inappropriate. The qualitative data were seemingly analyzed appropriately but there was little detail on how the analyses progressed. Mostly emergent themes were presented.

The data collection methods included sampling considerations and instrumentation. The samples were often convenience samples; meaning people who were easy to obtain data from. This often included selecting the best cases or ones that were geographically near by. Response rates were often low and non respondent studies were rare. Most of the instruments that were provided appeared to be sound but little information on the construction of the instruments or indications of their validity was provided. One exception to this was a student assessment instrument but the analyses provided showed that it was probably not a particularly strong instrument. There were many instances of

case studies and interviews with varying amounts of detail about how these were conducted.

The last consideration is the design of the evaluations. As suggested above, an overarching formative question would be “is the program working in the expected way” and the two major questions for summative evaluation would be “was there change” and “was the change more than in some other condition.” All of the evaluations addressed the formative question in some way. They all reported on how the program was operating and how that operation fit within NASA goals. They all also provided recommendations as to how the program might be improved or changed. Most also provided a good deal of information about how the participants in, and deliverers of, the programs felt about the program. Overall they provided very interesting descriptive information about the programs from the perspectives of those involved in them. The perceptions were overwhelmingly positive. In terms of the summative questions there were very few evaluations that actually looked for pre to post change. However, there were several retrospective questions that asked participants to comment on how much they felt they had changed. Again most people felt that the programs had affected them very positively. There were only a very few small attempts at comparative studies and these were flawed by selection bias.

It appeared from reading the reports that few if any of NASA decisions about educational programming was based on the evaluation reports. Three possible reasons for that come to mind. First, the evaluations may not have provided the information needed to make decisions. Second, the political environment moves much more rapidly than the evaluation environment and perhaps the reports were not available when decisions needed to be made. A third reason may be the very broad goals that are specified for the NASA programs. It would be very hard for any program, much less one with the limited funding available for NASA programs, to achieve these goals in any depth. It may be that these issues might be resolved with more careful discussions when planning evaluations. The Planning and Evaluation section recommends that evaluations be carefully negotiated. Perhaps with discussion more targeted goals could be determined that would be more amenable to evaluation.