Using Evaluation to Increase and Measure the Impact of Education: Evaluation within SMD

Hilarie B. Davis, Ed.D.

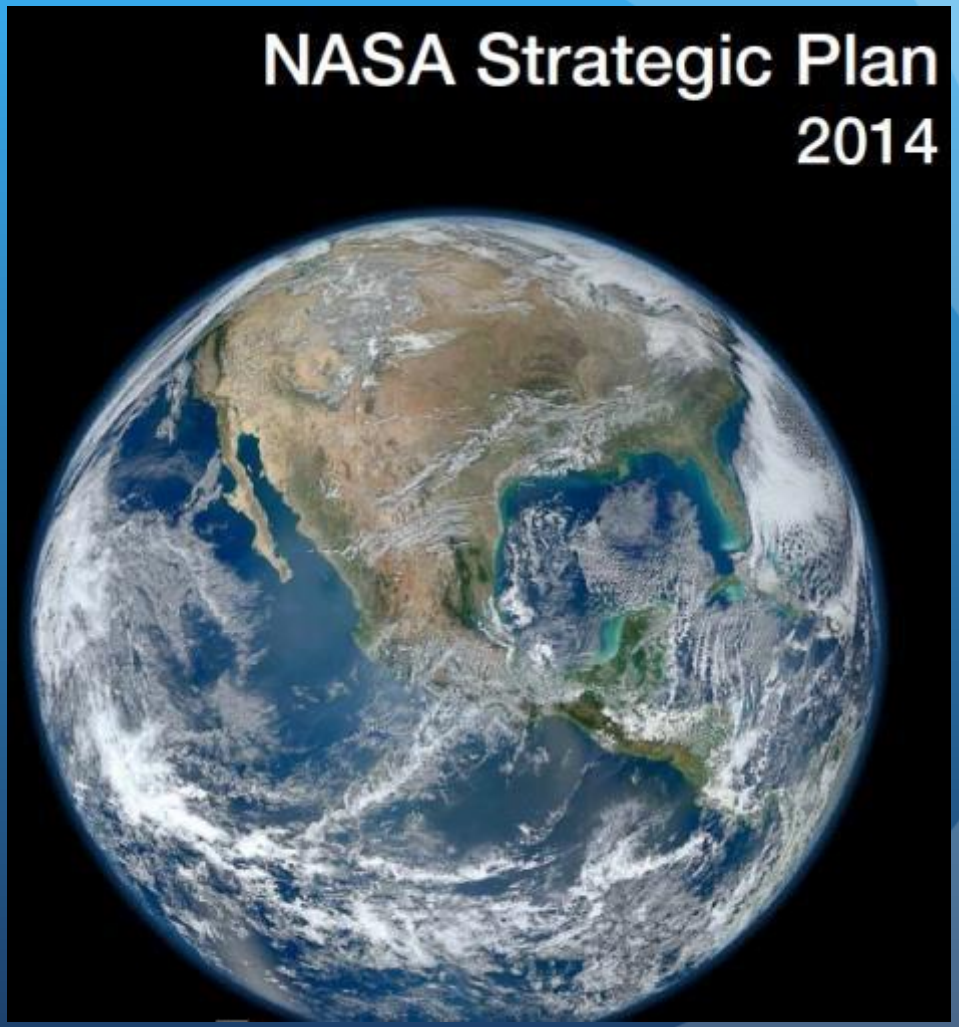
TLC Inc.

Guiding questions for our session

- What are the barriers to evaluation? What strategies have been used or recommended for addressing these barriers?
- How can we know NASA makes a difference in STEM education?
- Why and how are NASA projects evaluated?
- What are examples of evidence that evaluations improves programs?

NASA's Education Mission

Advance the nation's STEM **education** and workforce pipeline by working collaboratively with other agencies to engage students, teachers, and faculty in NASA's missions and unique assets



It’s the year 2020…

Evaluation

of the social contract of SMD education with the country

is required, funded, and used for improving outcomes

Evaluators

have constructed explanations of what works

and built models for effective evaluation

that are used for funding decisions

Evaluation questions have been addressed

*What is **NASA's unique contribution** to education? (formal and informal)*

*What interests **students** about NASA? How and under what conditions does their interest in NASA affect their interest in STEM over time?*

*What do **teachers** want and need from NASA to accomplish their objectives? How do they access these resources? Why and how do teachers use NASA content and missions? With what results?*

*What do **out-of-school programs** need from NASA to accomplish their mission? What resources do they need? What results do they get, under what conditions?*

All this came about in 2015 when barriers to evaluation and strategies for overcoming them were identified

Barriers to Evaluation

From **surveys of NASA education specialists**:

- Seems like something done by someone else for someone else instead of for improving the program
- Isn't close enough to the work being done to be meaningful
- Isn't realistic in its scope or methods
- Is too costly for the perceived value
- Feels like an audit or judgment of the people and/or program

Evidence showed that the barriers could be overcome

- Over 200 people attended evaluation sessions offered through SMD forums
- Attendees at the evaluation sessions reported it had significantly affected their understanding of evaluation, their perception of its value, and their intention to use it in the future
- Astrobiology Institute educators involved in ongoing professional consultations with an evaluator embedded evaluation in one or more of their projects that resulted in increased impact

Evaluation Strategies for Overcoming Barriers Were Identified

- Embed evaluation in the whole project cycle – provide feedback and support for this
- Give the evaluation credibility by involving the stakeholders appropriately
- Build the evaluation around questions that are important
- Use reasonable, practical approaches to collect data
- Be clear about the purpose of the evaluation
- Use the results of the evaluation to guide decision-making about program elements, goals, and funding

Measured IMPACT pre 2015

GPM Lessons Make a Difference

Does teaching curriculum concepts through the Global Precipitation Mission increase student knowledge as well as, or better than the standard curriculum?

YES

GPM students did as well as county curriculum students oncounty knowledge tests for seven lessons during the year (N=250)

GPM students did better than county curriculum students on retention of concepts in an end of year test (p=<.01)

http://pmm.nasa.gov/GPM

Intern Program Results Improved through Needs Assessment

Rural high school students worked with Ames astrobiologists to study extremophiles in nearby Lassen Volcanic National Park. Before they started and 5 other times during the year, students answered core questions about astrobiology that the science team used to guide their interactions with students. Students collected and analyzed data and presented their findings to the community, demonstrating their understanding of the science. All the interns report they intend to continue studying some area of biology.

http://astrobiology.nasa.gov/nai/reports/annual-reports/2011/arc/epo/astrobiology-student-intern-program-at-lassen-volcanic-national-park/

Sun-Earth Venus Transit Inspires Millions

How does NASA bring an astronomical event to teachers, students and the public? What is the reach?

The SED team led a year long preparation effort with astronomers and educators from around the world, provided a website of multimedia resources, supported venues offering events, webcast for six hours during the transit, and put out information on social media before, during and after the event. Evidence includes: >1 million events worldwide, 380 million web hits, 28 million twitter impressions, 2.5 million facebook impressions and 7.8 million shares, 7.7 million webcast streams, 30,000 mission packets, 4300 museums, 13000 educators

http://sunearthday.nasa.gov/transitofvenus/

MMS Provides Models for Classroom Implementation

Teachers used MMS to teach students how to build models and use them to understand engineering design, the mission's purpose, and instrumentation that will return data. The Magnetosphere Multiscale Mission provided paper, card and full size resources on models along with video and audio information on the mission. Students' models were accurate and their reflections showed an understanding of the mission and the science behind it.

http://mms.gsfc.nasa.gov/

Heliophysics Education Ambassadors Get Results

Can teachers develop an understanding of multiple missions, the science involved, and the data being collected to use in their own classes and teach other teachers?

YES

In a summer workshop, 84 HEAs heard about NASA missions from education specialists, developed viable, accurate lesson plans with specialist support, used them with their students, and did effective workshops for other teachers (4000 teachers) who taught their students (400,000).

http://cse.ssl.berkeley.edu/hea/

Adler IBEX After School Club Increases Student Knowledge and Interest

Can an informal education program affect students' knowledge and interests?

YES

Middle school students showed statistically significant increases in their knowledge of space science concepts after one semester attending an after school club. They reported increased interest in science and confidence in learning science.
http://www.smdepo.org/project/5696

Standards for rigor were established

- Impact was concretely defined
- Diagnostic information on current programs
- A clear path to improvement through feedback
- A rubric to guide development of new programs
- Professional consultations with an evaluator to ensure value and practicality of evaluation

Impact has been defined concretely

The intended and unintended effects on the Behavior, Attitudes, Skills, Interests, and/or Knowledge (BASIK) of participants.

Impact is determined based on the *data you collect as evidence of impact (the results)* and the *rigor of the methods and measures* you use to collect those data.

Results
What are the data saying about how well the objective was met?

X

Rigor
How confident can we be about the data based on how they were collected?

Evaluation is Embedded Throughout the Project Cycle to Increase and Measure Impact

Needs Assessment	Determine the context for impact
Objectives	Define specific impacts
Design	Create plan to achieve impact
Implementation	Deliver the design to achieve impact
Outcomes Assessment	Measure the impact

Impact Categories from NSF

- **B** Behavior
- **A** Attitude, aspirations
- **S** Skills
- **I** Interest, engagement
- **K** Knowledge

Objectives

SPECIFIC

MEASURABLE

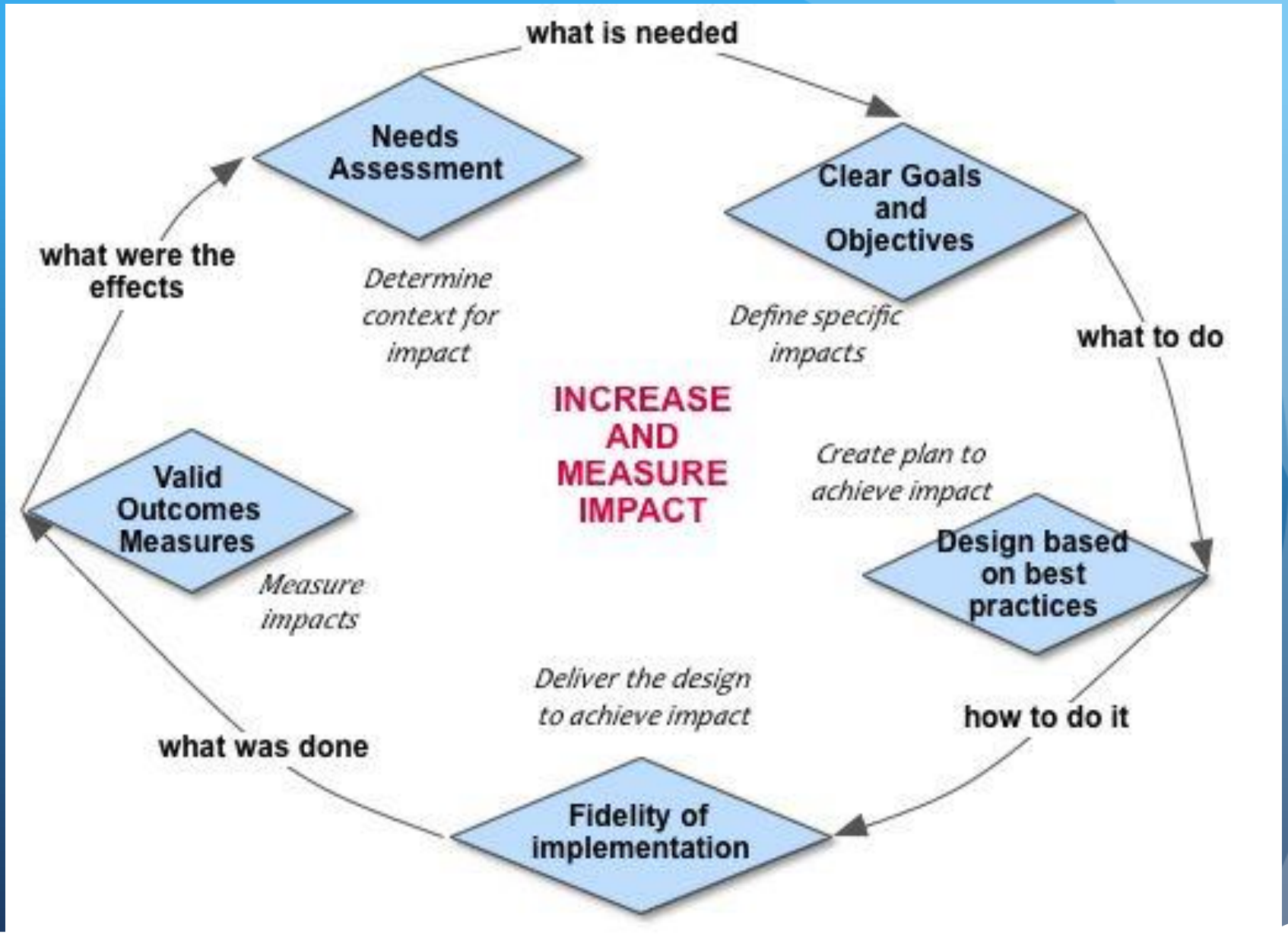
ACTION-ORIENTED

REALISTIC

TIMELY

SMART

Evaluation is embedded in the project cycle



A rubric shows a clear path to success

Needs Assessment

Fair (1)	Good (2)	Very Good (3)	Excellent (4)
Prior experience; "Seems like a good idea"	Research on what works; Literature review on similar programs/ products/populations / goals	Conversation with and/or direction from stakeholders (Focus Group); Experts review the ideas/plan	Survey of or pilot with potential audience/users about the draft program

Goals and Objectives

Fair (1)	Good (2)	Very Good (3)	Excellent (4)
General direction; Understood by team; Agenda substituting for objectives	Explicit, written; For a target audience	Objectives are SMART*: Specific, Measurable, Action-oriented, Realistic, Time- bound	Logic model of inputs, outputs, and outcomes in place

A rubric shows a clear path to success

Design: *How evidence- or research-based is the design?*

Fair (1)	Good (2)	Very Good (3)	Excellent (4)
Series of activities; Uses what has worked before	Based on objectives; Connects to standards; Includes contingency plans for emerging needs	Thematic; Has continuity; Participatory, personalized, responsive; Uses advanced organizers	Developmental; Embeds evaluation/ reflection

Implementation: *How true to the design is the implementation?*

Fair (1)	Good (2)	Very Good (3)	Excellent (4)
Facilitators prepare to implement the design	Collect and use feedback during implementation	High fidelity to design OR implements contingency plans to meet objectives if needed	Participants able to monitor their own progress against objectives

A rubric shows a clear path to success

Outcomes: *What is the evidence of impact on BASIK?*

Fair (1)	Good (2)	Very Good (3)	Excellent (4)
Post only survey or reflection; Follow up survey or interview; Web stats; Anecdotes; Facilitator reports	External evaluator observes, or does case studies; Pre/post self-report survey, reflections; Post only measure (test, retrospective survey, task)	Pre/post measures (tests, performance tasks, observation); Pre/post follow-up	Comparison group studies (quasi-experimental); Experimental studies (random assignment)

Feedback from NASA education specialists about evaluation
has shown significant changes in their mental models of evaluation
and
use of evaluation in their practice

eVALUEuate

It is clear that evaluation is, to a certain degree, as important as the work itself. When possible, we will include an evaluator right at the beginning as we design new projects.

There is no point in doing the work unless you can prove its worth.

Evalicity

Taking a hard, honest look at how our programs really measured up in terms of rigorous evaluation was very insightful! Now I wouldn't want it any other way!

The evaluation consultation on the activities allowed us to critically look at our programs and determine what we need to do to improve them and get the results we desire.

Evalability

We were very locked into the notion that evaluation means people filling out surveys. It was refreshing and helpful to brainstorm many other ideas and techniques for gathering data.

Having a rubric to consult is incredibly helpful. It's great to have a 'recipe' to follow that will improve the quality of our work.

Evalboration

Exchanging ideas with colleagues and working through difficult scenarios (such as how to evaluate a booth at a public event) has been incredible and in the process we have learned many new methods and processes that will improve our work.

Why have these strategies worked?

- People like feedback – not judgment. Judgment feels punitive while feedback feels helpful
- People want to do well – they set out to succeed, not to fail – so they appreciate a fair assessment that may help them improve
- Evaluation throughout the project cycle improves it every step of the way so there are lots of chances to improve
- People want answers to their questions, so when they help develop the questions, they care about the answers

Why have these strategies worked?

- People improve when they have a clear path to getting better, which is why they say the Project Cycle Rubric helps
- People delivering programs know where and how good data can be most effectively collected
- Evaluators do a better job when stakeholders evaluate their evaluation plans, methods and measures for value and validity. Stakeholders are also experts
- Decisions based on good data about a program are honest and productive; decisions made without good evaluation data are suspect and feel arbitrary, which discourages productivity

In 2020 things people can count on:

- Valid evaluation data is valued by implementers, managers, and organizations to offer, improve, and make decisions about what programs to fund
- Evaluation is systematic and supported
- Evaluation is embedded to promote improvement in all stages, so increases the impact
- Stakeholders are involved collaboratively in evaluation, making it is more valuable to them and more rigorous because of their involvement and interest in the evidence

In 2020 things people can count in SMD:

- All program descriptions, evaluation plans, evaluation reports are in an online database (begun in smdepo.org)
- Teachers have an online NASA identity - membership in the NASA community that shows resources used, interests, activities – over time and cross program
- Students have an online NASA identity - membership in the community that shows participation, ideas, interests, and activities – over time for longitudinal tracking
- Teachers have contact with NASA professionals (educators and scientists) as part of the community so they feel close to the premier research effort of our time

NASA is the premier
research endeavor of our time

Evaluating NASA's education efforts
parallels its scientific efforts

Evaluation examines
the how and why
as well as
the what and that

Through evaluation
we are able to collect evidence
and
develop explanatory models
of
how to bring back the wonder
for teachers' and students'
to
know, care about, and pursue
NASA and STEM learning

Panel

Frances Lawrenz, University of Minnesota Lawrenz@umn.edu

- Barriers in the past
  - Underfunded-tension of program and evaluation, unrealistic expectations, unable to address 'real' questions, sampling bias
  - Time-short tem or retrospective, participant opinion
- Future (Acknowledge values—utilitarian-pluralistic)
  - Recent models—Developmental evaluation (principles), Educative values engaged evaluation, culturally relevant evaluation
  - Two related—Utilization focused (decision making) and Context and Input from CIPP rather than Process and Product
  - Less is more Pick your battles

Jenny Gutbezahl Brandeis University
jgutbeza@brandeis.edu

Helped lead the evaluation of the prior NASA Space Science E/PO effort (1997-2007)

Am seeing many of the same challenges we saw then:

- Culture clash between scientists and educators
- Lack of coordination across the system, leading to gaps and redundancies
- Challenges between going for depth vs. breadth

Am seeing many of the same strategies to address these problems:

- Creating common goals to overcome culture differences
- Going to the users to discover their needs

Jenny continued …

Current evaluation places more emphasis on "empirical evidence”

- By which they mean numbers, which really are not any more empirical than qualitative data
- Leads to an emphasis on breadth, because its easier to count noses than measure true impact
- Leads to a challenge similar to what educators face

In the future:

- Want to create meaningful experiences for learners
- Need to meet standards that may not be aligned with educators' understanding of best practices

Bonnie Eisenhamer bonnie@stsci.edu
Space Telescope Science Institute

- A clear understanding of program purpose and goals, and evaluation questions that are feasible, appropriate to the program, and address stakeholder needs
- Evaluation questions are well matched to define purpose and strategies
- Evaluation questions and methods are appropriate to the stage / maturity of the program
- Front end planning is needed, plan evaluation with the end outcomes in mind
- A new direction for SMD education will require the involvement of an evaluator from program inception