What and How Are We Evaluating? Meta-Evaluation of Climate Education Projects Funded by NASA

Ann M. Martin, Lin H. Chambers, and Margaret R. Pippin
NASA Langley Research Center, 21 Langley Boulevard, Hampton, Virginia 23681, USA

Abstract. NASA Innovations in Climate Education (NICE) at Langley Research Center has funded 71 climate education initiatives over four years, each evaluated separately by external evaluators. NICE has undertaken a systematic meta-evaluation, seeking to understand the range of evaluations, approaches, and methods represented in this portfolio. When NASA asks for evaluation of funded projects, what happens? Which questions are asked and answered, using which tools? To what extent do the evaluations meet the needs of projects and program officers? How do they contribute to best practices in (climate) science education? These questions are important to ask about general STEM education work; the NICE portfolio provides a broad test case for thinking strategically, critically, and progressively about evaluation in our community. Our findings can inform the NASA, ASP, and STEM EPO communities and prompt us to consider a broad range of informative evaluation options.

1. Project Overview

NASA Innovations in Climate Education (NICE) is a funding opportunity for climate change education initiatives in K–12 and higher education contexts. Since 2008, NICE (formerly Global Climate Change Education/GCCE and Innovations in Global Climate Change Education/IGCCE) has funded 71 projects across the United States. The portfolio includes research experiences for students and teachers, professional development opportunities, and the development of data access and modeling tools.

Through funding and strategic project management activities, NICE pursues and contributes to the following goals: (1) increasing climate literacy and engagement of public audiences; (2) contributing to the creation of the future workforce in climate-related STEM fields; and (3) developing successful resources, experiences, and activities that connect NASA climate-related science to pre- and in-service teachers/faculty and their students.

Each of the funded projects proposes and carries out its own evaluation plan, in collaboration with external or internal evaluation experts. Therefore, projects carry out a broad range of evaluations, incorporating various methods and approaches. NASA’s evaluation requirements have varied somewhat with each funding solicitation, but generally request evaluation plans that are “appropriate” for the proposed intervention and that discuss “data collection and analysis procedures.”

Because of the breadth and size of the NICE portfolio, these evaluation plans can serve as an interesting test case for understanding the range of evaluations that may be conducted using NASA funding to external awardees. In this paper, the authors
Martin, Chambers, and Pippin share their findings from a pilot meta-evaluation study of the evaluation plans within the NICE portfolio. This pilot assessed evaluation plans from a randomly selected 35 of the 71 projects. Findings from the pilot prompted further, and new, questions, which will be explored in the full study using all 71 evaluations.

2. Why Meta-Evaluation?

Meta-evaluation is, most simply, the evaluation of evaluations; it provides a glimpse into, and a check on, evaluations that may serve a variety of goals and interests. Evaluation practitioner and theorist Daniel Stufflebeam has argued (2001) that meta-evaluation is a *critical* component of a strong culture of evaluation, and that attention to meta-evaluation is in the interests of both evaluators and those who fund projects.

A meta-evaluative study can provide a baseline understanding of the kinds of evaluations carried out within the specific context of a case like NICE. These explorations can help the community think about evaluation strategically, critically, and progressively. This pilot study is primarily descriptive in nature and was designed to provide basic information about evaluation in the NICE portfolio: When NASA asks for evaluation of funded projects, what happens? Which questions are asked and answered, using which tools?

3. Methods and Findings from a Pilot Study of 35 NICE Evaluation Plans

Descriptive meta-evaluation criteria were developed using the Program Evaluation Standards (Yarbrough et al. 2011) and existing meta-evaluation checklists (e.g., Scriven 2012). Additional evaluation characteristics of interest were identified from the NICE logic model and the management team’s experiences. A pilot rubric was developed to identify key descriptive characteristics of interest; the rubric and list of characteristics will be revised and expanded for the full study. Each evaluation plan was read, coded, and matched to each element of the rubric by a single coder. Here, we will focus on findings in three key areas: (1) proposed project outcomes, (2) evaluation outcomes emphasis, and (3) evaluation design.

3.1. Proposed Project Outcomes

First, we investigated the outcomes and goals that projects identified as important, which would presumably correlate tightly to the inquiry conducted in their evaluations. From the 35 projects, we identified a total of 129 explicitly stated project outcomes. We then considered the temporality of these outcomes (were they short, intermediate, or long-term goals?) and the type. Types were adapted from Friedman and the NSF’s (2008) definitions of outcomes focused on Knowledge/Awareness, Engagement/Interest, Attitudes, Behaviors, Skills/Competencies, and Other.

When considering temporality, we defined short-term outcomes as those that would take place early in the project implementation, intermediate-term outcomes as those that would likely take place by the end of the funding period, and long-term outcomes (what could be referred to as “impacts”) as those that would likely take place after the end of the funding period. Figure 1 displays our finding that the majority of explicitly stated goals were focused on intermediate-term outcomes.
When we investigated the type of outcome, several interesting characteristics of the portfolio of evaluations emerged (see Figure 2). First, we found that the Friedman/NSF outcome types were generally sufficient for describing the portfolio, but that several other dimensions emerged that were important for our projects. We therefore elected to make separate categories to account for objectives dealing with deliverables, career choices (primarily applicable to projects working with students), and diversity (particularly applicable to 2011 NICE awardees, who are funded through the NASA Minority University Research and Education Program, MUREP). In fact, one of these categories, deliverables, was the most common.

Second, this analysis determined that, of the five Friedman/NSF outcome categories, two were very apparent in the noted project goals: Knowledge/Awareness and Skills/Competencies. Given the earlier-stated objectives of the NICE funding program, it is not surprising that there is a great deal of emphasis on these categories. On the other hand, given the focus on educators and resources to support their work in the classroom, the limited focus on Behaviors is notable. The same is true of the limited emphasis on student outcomes related to Careers and Engagement/Interest.

### 3.2. Evaluation Outcomes Emphasis and Data Sources

Though each project identified goals and outcomes, overall, the evaluation plans tended to focus on a particular, limited set for intensive inquiry. For each evaluation plan, we assessed the extent (minimal, moderate, or strong) to which it focused on each of the five outcome types. This flexible approach allowed us to capture evaluation plans that considered more than one outcome area.

As demonstrated in Figure 3, below, the Knowledge/Awareness outcome type was most common. Skills/Competencies received little attention, although the project outcomes had a high concentration in this type. Experience in STEM education evaluation suggests that this may be due to the prominence of content-knowledge assessments in educational research and evaluation.
An investigation into the data sources used across the portfolio of evaluations may support this interpretation. Once again, we assessed the extent (minimal, moderate, or strong) to which each evaluation focused on a set of data source types. The rubric included Assessments (e.g., exams or surveys), Interviews/Focus Groups, Observation, and Document Analysis, as well as a more general Self-Report category (e.g., reflections or journals where participants self-assess). This data is displayed below in Figure 4: note that Assessments and Self-Report sources of data are heavily emphasized in NICE evaluation plans. Sources of data that are more removed from participants’ perceptions, such as Observations (e.g., in the classroom) or Document Analysis (e.g., external reviews of lesson plans) were less common.

![Figure 2. Project outcomes organized by type.](image)

3.3. Evaluation Design

Finally, we considered the fundamental design of evaluation plans within the NICE portfolio. We identified a primary research design for each plan (with the exception of some evaluation plans where the design was unclear). While it would have been surprising to see a significant number of evaluations using a true experimental design, this extended even to quasi-experimental designs (see Figure 6). Nonexperimental designs were by far the most common, though these designs can suffer from significant validity threats, as they do not include comparison groups. One evaluation design included a true experiment.

In addition, we found that all evaluation plans were to some degree mixed-method (combining qualitative and quantitative data collection and analysis). The majority of evaluation plans did not identify data analysis techniques that would be applied to qualitative (29 of 35 plans) or quantitative (24 of 35 plans) data.
4. Conclusions

Evaluations within the context of a portfolio like NICE are constrained in a variety of ways, by limited budgets and resources, and a broadly interpretable request for evaluations within the funding solicitation. For instance, fully experimental evaluation designs are costly, resource-intensive, and carry serious ethical concerns. Such an approach would thus be, appropriately, rare in a context like NICE where projects last
up to three years and have moderated costs (around a few $100,000). In response, a
community of STEM education evaluators has developed a “culture” of informal, fluid
guidelines for what is typical and normative for this kind of evaluation.

This context is reflected in our preliminary findings from the pilot study which
indicate that the evaluations within the NICE portfolio strongly focus on intermediate-
term outcomes, summative evaluations, and nonexperimental designs, while leveraging
assessments, interviews, and self-report data as sources of information. By contrast, the
evaluations largely do not take into account long-term outcomes (e.g., impact) or observ-
utional data sources, and front-end, quasi-experimental, and experimental evaluation
approaches are rare.

Carrying out, and then sharing, this pilot study has prompted us to consider a
range of further questions that can be included in the full study. For instance, we plan
to determine the extent to which each project’s listed, purported objectives are explicit-
ly and truly reflected in the evaluation plan’s proposed questions. We anticipate that
these results may help identify some of the current norms in the evaluation of federally,
competitively awarded, small-scale STEM education initiatives.

Acknowledgments. AMM was supported by an appointment to the NASA Post-
doctoral Program at NASA Langley Research Center, administered by Oak Ridge As-
associated Universities through a contract with NASA.

References
Projects (http://caise.insci.org)
(http://michaelscriven.info)