

## Learning the Ropes of Program Evaluation: A Case Study

By Linda Urban, Marnie Christenson, and Susan Benson

*Tales from the Field, a monthly column, consists of reports of evidence-based performance improvement practice and advice, presented by graduate students, alumni, and faculty of Boise State University's Organizational Performance and Workplace Learning department.*

### Introduction

In the day-to-day work of many professionals, it can be challenging to find the time to delve deeply into new processes, try them out, and get the feedback necessary to build skills and capacity. One of the benefits and luxuries of returning to school as a graduate student is getting the opportunity to do just that.

As graduate students in the Boise State's Evaluation Methodologies class, we were in the middle of two worlds. The class would teach us the principles and concepts of conducting an evaluation under the guidance and direction of our professor. At the same time, we took on the role of evaluators for a real client, who needed credible and valuable results from our evaluation project.

In this article, we tell the story of conducting a program evaluation as an external evaluation team. At the same time, we want to share some key elements of the program evaluation framework we followed—one that was new to each of us, and is now an important addition to our toolkits. This framework can be applied to both simple and more complex evaluation projects.

We worked with a rich set of tools that complemented each other and added depth to the evaluation process: Michael Scriven's (2011) Key Evaluation Checklist (KEC), a Program Logic Model (W.K. Kellogg Foundation, 2004), Brinkerhoff's (2006) Training Impact Model, and several tools from the graduate course to consolidate and triangulate our data.

### Background and Evaluation Request

The program we evaluated was a Graduate Student Retreat program provided by Transdisciplinary Work in Climate and Agriculture (TWCA) (a pseudonym). TWCA is a five-year research project comprised of an interdisciplinary group of scientists and other researchers from several different universities in the United States. Graduate students are an integral part of the TWCA project. They contribute to research for TWCA at the same time that they take coursework in their area of study and work on research for their own thesis or dissertation.

In year two of the TWCA project, they held a weekend Graduate Student Retreat at which the geographically dispersed graduate students could connect face-to-face. It also provided an opportunity for key TWCA leaders to communicate expectations of the graduate students, provide information about TWCA, and help the students see how their work fit into the overall project goals.

The main goal of the Graduate Education Component of the TWCA project is to produce a group of high caliber scientists who have the tools they need to do research in their discipline, plus additional tools that allow them to a) work collaboratively across disciplines, and b) communicate effectively with stakeholders and educators. Our clients for this project, the TWCA Education Team Lead and the Education Coordinator, wanted to know whether the retreat helped the project move toward that goal, and whether it met the needs of the graduate students.

## Evaluation Purpose and Questions

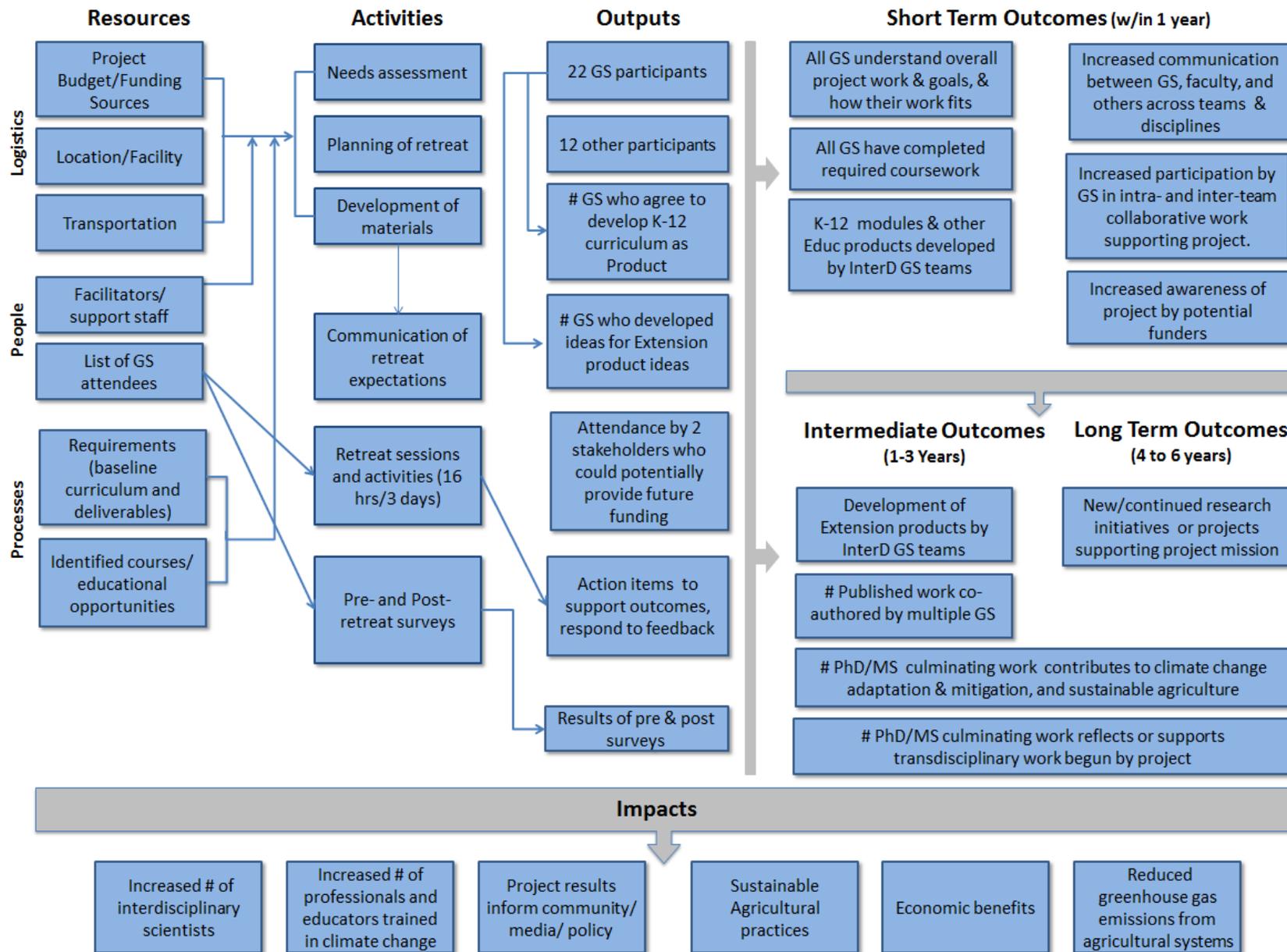
One of the first things we had to decide was whether to do a formative or summative evaluation. We decided it was possible and important to do both. Our clients wanted both an assessment of the worth of the event itself, and wanted to know where they could improve.

We conducted this evaluation about two months after the retreat, and focused on two main evaluation questions:

1. How successful was the Graduate Student Retreat in moving toward the goals of the Graduate Education Component of the TWCA project?
2. How much progress were the graduate students making toward both short-term and long term goals that the project had for them?

## Diving into Detail and Seeing the Big Picture: The Retreat in Context

To understand the graduate student retreat and how it fit into the overall goals of the project, we worked with our clients to develop two models: A Program Logic Model (Kellogg, 2004) (Figure 1) and an Impact Model, based on Brinkerhoff (2006) (not shown). Developing both models was helpful. The Logic Model showed the bigger picture, including the resources and activities that went into the event, the outputs from it, and the desired short-term and longer-term outcomes and impacts. The Impact Model provided more detail about the knowledge, skills, and behaviors that graduate students were expected to acquire. These models continued to be important tools as we moved forward, especially as we worked with our clients to identify what criteria to evaluate.



GS=Graduate Student

**Figure 1. Graduate Student Retreat Logic Model.**

## Evaluation Components

In our class, we discussed components of an evaluation project. We worked through each of the following as part of our evaluation process:

- Establishing evaluative dimensions (*what should we assess?*)
- Determining importance weighting among dimensions (*which dimensions are more important than others and why?*)
- Developing instruments and collecting data (*what data would help answer the evaluation questions? How could we gather the appropriate data?*)
- Constructing rubrics (*the rules used to assign ratings*)
- Measuring performance against standards/rubrics (*applying the rubrics to the data we gathered*)
- Synthesizing and integrating evidence into a final conclusion (*making sense of it all*)

### Establishing Evaluative Dimensions and Determining Importance Weighting

Using the logic model and impact model to guide our discussion, we worked closely with our clients to identify which dimensions of the Retreat to evaluate —the outcomes that really mattered to them. We settled on short-term outcomes that could be assessed in the two months following the retreat, and ones that our clients could not readily assess without this evaluation.

We selected dimensions related to both “process” (the content and implementation of the event itself) and “outcomes” (impacts on immediate recipients and other impactees), and designed questions related to each. We then asked a selection of stakeholders to weigh each dimension’s importance. We used this weighting later, when we applied rubrics to evaluate the success of the Retreat. Table 1 displays the criteria, questions, and weightings.

Although most of the questions were closely linked to project goals, our clients were also interested in learning about other outcomes that may have resulted from the retreat. As a result, we included a “goal-free” question with a dimension of “other outcomes.” That brought the number of criteria to seven, which was manageable and allowed us to evaluate the program from a systemic perspective.

**Table 1. Evaluative Dimensions and Dimensional Evaluation Questions**

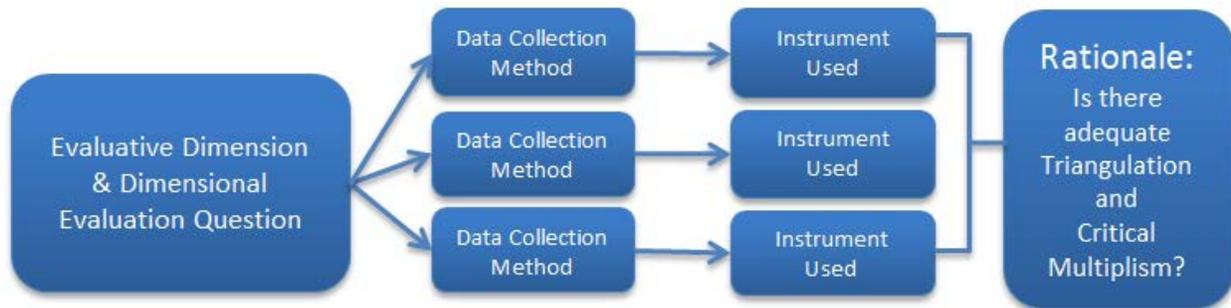
	Dimension	Question	Weighting
Process	Content/Design	Was the retreat designed to provide activities and information that the graduate students needed?	Very Important
	Implementation (Delivery)	Were the activities at the event conducted in a way that supported the goals for the retreat (specified in the Content/Design question)?	Very Important
Outcomes	Project Knowledge and Awareness	Do graduate students have a broad understanding of TWCA project work and goals, and know where their work fits in?	Very Important
	Communication	Has there been an increase in communication between grad students, faculty, and other researchers across teams and disciplines?	Important
	Collaboration	Has there been an increase in participation by graduate students in collaborative work supporting the TWCA project (both within teams and across teams and disciplines)?	Important
	Improved Support	Has the Education Team improved their support of graduate students by acting on needs identified during the retreat?	Very Important
	Other Outcomes	What other outcomes resulted from the retreat (from both the participants' and the organizer's points of view)?	N/A

## Collecting Data

An important part of any evaluation is gathering data that relate to the evaluation questions. In order to establish the data’s validity, they should come from multiple and different types of sources using varied methods. Technically, this is called using critical multiplism and triangulating the data to confirm or disconfirm the findings (Davidson, 2005).

We used a number of data collection methods. We surveyed and interviewed graduate students, faculty advisors and Education Team members, reviewed retreat materials and graduate student support materials, observed related meetings and reviewed observation notes from the retreat. As we developed the surveys, we aligned questions to each of the identified dimensions to make sure we were asking the questions that would provide us with relevant data.

We created a table that linked our data collection methods to the evaluation dimensions, and helped us think through and articulate whether our methods provided adequate triangulation and critical multiplism. Figure 2 shows the flow of our thinking, and Table 2 shows an excerpt from our actual data collection strategy.



**Figure 2. Data Collection Strategy Aligned to Evaluation Dimensions.**

**Table 2. Excerpt of Data Collection Strategy Table**

Category	Evaluative Dimension & Specific evaluation question	Data collection method	Instrument used	Rationale for using triangulation & critical multiplism
<b>Process</b>	<b>Content/Design:</b> Was the retreat designed to provide activities and information the grad students needed, in order to: <ul style="list-style-type: none"> <li>Understand the overall TWCA project</li> <li>Know what is expected of them,</li> <li>Establish interdisciplinary teams to work on “product” development, and</li> <li>Lay a foundation for building capacity as a</li> </ul>	1. Review of retreat agenda and planning notes	1-1.Document review checklist	Input from those who planned the retreat was triangulated with perceptions of those who were served by the retreat –the graduate students who attended it, one of the faculty advisors, and another person on the TWCA staff.
		2. Semi-structured interview with retreat planners about goals, informal needs assessment	1-2. Semi-structured telephone interview instrument (Retreat Planners)	
		3. Survey sent to 22 graduate students (11 respondents) who participated in the retreat	1-3. Web-based survey administered in Qualtrics	
		4. In-depth telephone interviews with selected grad students	1-4. Semi-structured telephone interview instrument (Grad Students)	The combination of surveys, interviews, and extant data reviewed

Category	Evaluative Dimension & Specific evaluation question	Data collection method	Instrument used	Rationale for using triangulation & critical multiplism
	group of interdisciplinary climate and agricultural researchers	5. Survey sent to 16 faculty (5 respondents) who advise the graduate students who attended	<b>1-5.</b> Web-based survey administered in Qualtrics	provided multiple perspectives and served to either reinforce our findings or identify areas of inconsistency related to whether the design of the retreat provided activities and information the grad students needed.
		6. Interview with one of the faculty advisors	<b>1-6.</b> Semi-structured telephone interview instrument (Faculty Advisors)	

We achieved a smaller survey return rate than we wished. We received survey responses from 50% of graduate students, 11 out of the total of 22 students who were involved in the retreat. And only five out of 16 faculty advisors (31%) responded to the survey. With hindsight, we would have consciously considered how to encourage responses, as part of our data gathering strategy.

### Constructing Rubrics and Using Them to Evaluate Performance

Data are essential for establishing evidence—but how do you put them together in such a way that you can assign a quality rating or judgment? That’s where rubrics come in. They specify the rules to follow when assigning ratings. We developed two rubrics for this project:

- **A triangulation rubric**, that we used to rate each dimension we were evaluating by triangulating the data we gathered from various sources.
- **A synthesis rubric**, that we used to determine the overall quality of the Graduate Student Retreat by synthesizing multiple dimensional ratings by their importance weightings.

We created these rubrics before analyzing the data to avoid bias in deciding how ratings would be determined.

#### *Triangulation Rubric*

Our stakeholders told us that the most important data sources were the Graduate Students, so we used their survey responses as primary data and other data sources as secondary or tertiary sources.

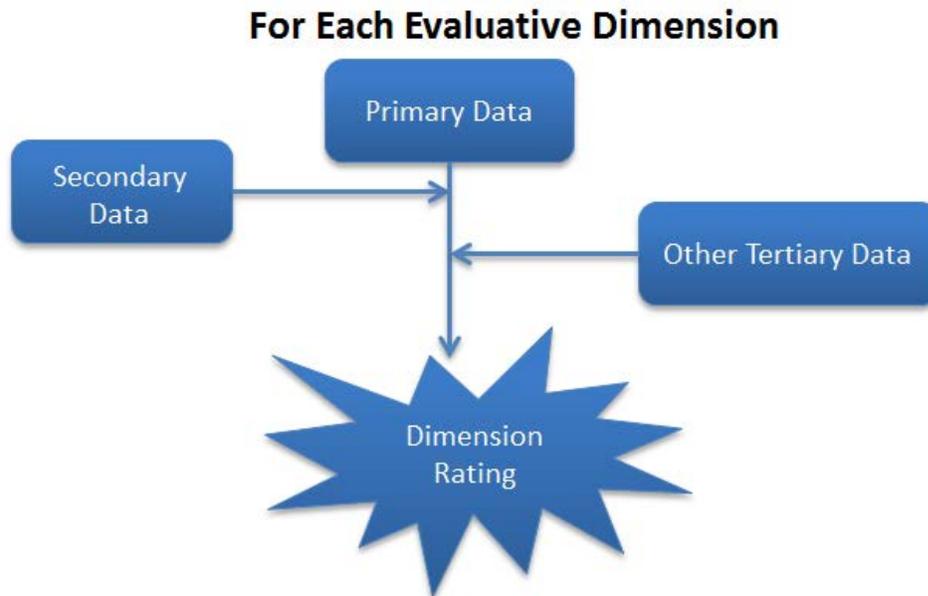
The survey scale was from 1 to 5, where 1=strongly disagree, and 5=strongly agree. We developed the following rubric scoring framework for the survey:

- *Excellent* – Survey average score is between 4.0 and 5.0
- *Good* – Survey average score is between 2.1 and 3.9
- *Poor* – Survey average score is between 1.0 and 2.0

We decided that we would make the decision about the quality rating for each dimension based on the Graduate Student survey data **as long as the secondary data seemed to support it**. We would consider upgrading or downgrading the quality for each category if the Graduate Student survey data result was close to the edge between two levels in the rubric (for example, if the average = 1.9 or 4.1), and/or if the **interview and other secondary data related to that dimension** strongly indicated against the survey results.

With seven dimensions and six data collection methods, we needed a concise way to assign ratings to each dimension. Using a table, we aligned each dimension with corresponding survey question scores,

interview answers, and other observations and document reviews. Looking at the data this way also helped us to know when to incorporate secondary and tertiary data as a “tiebreaker.” Figure 3 shows how we consolidated data to assign an overall rating for each category.



**Figure 3. How to combine data was part of our rubric.**

**Synthesis Rubric**

We developed a Synthesis Rubric (Table 3) to determine the overall quality of the Graduate Student Retreat. Taking the weightings designated by stakeholders into account (Table 1), the synthesis rubric specified what combination of ratings per dimension would indicate an Excellent, Good, or Poor overall quality rating for the retreat.

**Table 3. Synthesis Rubric**

Retreat Rating	Dimension Weighting	
	Very Important	Important
Excellent	All four very important dimensions rated excellent	Both important dimensions rated good or excellent
Good	All four very important dimensions rated good or excellent	One important dimension rated poor
Poor	Any of the four very important dimensions rated poor	One or both important dimensions rated poor

**Synthesizing and Integrating Evidence into a Final Conclusion – and Making Sense of It**

We consolidated the results of both rubrics into a table, shown in Table 4. Based on the synthesis rubric, the overall evaluation for the Graduate Student Retreat was “Good.”

**Table 4. Graduate Student Retreat Dimensions and Weighting**

Dimension	Graduate Student Retreat Overall Quality: GOOD			Weighting
	Excellent	Good	Poor	
Content/Design	✓			Very Important
Implementation	✓			Very Important
Project Knowledge and Awareness		✓		Very Important
Communication		✓		Important
Collaboration		✓		Important
Improved Support		✓		Very Important
	Excellent	Good	Poor	

Now we had ratings for each dimension, and for the overall retreat—a snapshot and starting point for discussion, but not enough information for our client to really see what specifically was valuable, and where actions could be taken to improve support and future events. What does “good” really mean?

The specifics of the underlying data indicated that the retreat acted as a catalyst in moving towards the goals of the Graduate Education Component of the TWCA project and helping graduate students make progress toward both short- and long-term goals. The data also showed that there were areas where our clients could improve support for graduate students, and where they needed further input from faculty in order to do so.

We provided a summary of strengths and recommendations for improvement for each category and supporting detail related to each category. Scriven’s KEC provided an excellent framework not only for conducting an evaluation, but also for building a report that provides multiple layers of detail to address interests and needs of different audiences. The course guidelines for aligning data gathering with evaluation dimensions and for building rubrics helped us gather the right data and organize them for both our analysis and reporting.

When we talked with our clients a year after the evaluation, we learned that they used our findings to enhance and flesh out a handbook for their students and that they referred to our report when planning their next retreat.

We think it is also worth noting that the act of evaluating something shines a light on it. We collaborated with our clients throughout; we talked about the needs of the graduate students, how they fit into the overall project, the goals for the retreat, the challenges inherent in supporting and integrating students on a complex project of this type. Our evaluation occurred at a time when the TWCA project was working on clarifying and communicating project goals for the graduate students. We hope our evaluation contributed to furthering that process.

## References

- Brinkerhoff, R. O. (2006). *Telling training’s story: Evaluation made simple, credible, and effective*. San Francisco: Berrett-Koehler.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.

Kellogg, W. K. (2004). *Logic model development guide*. Michigan: WK Kellogg Foundation. Retrieved from <http://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide>

Scriven, M. (2011). *Key evaluation checklist (KEC)*. Retrieved from [http://michaelscriven.info/images/KEC\\_4.18.2011.pdf](http://michaelscriven.info/images/KEC_4.18.2011.pdf)

### Author Bios:



Linda Urban is a long-time independent consultant and instructor who specializes in change management, needs assessment, evaluation, and performance improvement solutions to help individuals, teams, and organizations work more effectively. Linda is a 2013 graduate of Boise State's Organizational Performance and Workplace Learning Master's program. She can be reached at [linda@lucommunications.com](mailto:linda@lucommunications.com).



Marnie Christenson has been a learning design and organizational performance professional for over 15 years, currently working at DIRECTV as Sr. Manager of Training Design and Delivery for their Field Services organization. Marnie is a 2013 graduate of Boise State University's Organizational Performance and Workplace Learning Master's program. She can be reached at [marniechristenson@u.boisestate.edu](mailto:marniechristenson@u.boisestate.edu).



Susan Benson is a Senior Instructional Designer at American Express. She has worked in the organizational performance field for over 25 years in various industries, including telecommunications, the military and financial services. Susan is an August 2013 graduate of Boise State's Instructional and Performance Technology Master's program. She can be reached at [swbenson@att.net](mailto:swbenson@att.net).