Author-approved manuscript for

Hayward, C. N., Archie, T., Daly, D., Weston, T. J., & Laursen, S. L. (2022). The cycle of inquiry: Building effective evaluation relationships to support continuous improvement of faculty development initiatives. In S. Linder, C. Lee, K. High (Eds.), *Handbook of STEM Faculty Development*, Information Age Publishing, Charlotte, NC.
Order info: https://www.infoagepub.com/products/Handbook-of-STEM-Faculty-Development

# The Cycle of Inquiry: Building Effective Evaluation Relationships to Support Continuous Improvement of Faculty Development Initiatives

Charles N. Hayward,[1] Tim Archie,[1] Devan Daly,[1] Timothy J. Weston,[2] & Sandra L. Laursen[1]

[1] Ethnography & Evaluation Research (E&ER)

[2] Alliance for Technology, Learning and Society (ATLAS)

University of Colorado Boulder

**Abstract**

This chapter discusses lessons learned from more than a decade of conducting evaluation research about faculty development (FD) initiatives. We collaborate with faculty developers who provide FD to support a shift from notions of evaluation as summative judgment to a formative role where we use data to provide coaching and support continuous improvement throughout. Using Patton's (2003) Utilization-Focused Evaluation, we discuss strategies we use with faculty developers to (1) co-create appropriate evaluation methods, (2) provide formative feedback in actionable ways, and (3) capture improvement over time as part of our summative evaluation results. Throughout, we share concrete examples of practices we use to build productive relationships founded on what one of our partners calls the "cycle of inquiry."

## Conceptual Framework

Much of the work devoted to faculty development comes from a research perspective. It provides models of teacher change or methods for measuring it (e.g., Giersch & McMartin, 2014; Guskey, 2000, 2002; Kennedy, 2016). Evaluation is sometimes seen as just a necessary component of projects that must be done to meet funder requirements. Utilization-Focused

Evaluation (UFE) (Patton, 2003), on the other hand, is based on "the premise that evaluations should be judged by their utility and actual use" (p. 223). The process begins with identifying a primary intended user(s) and primary intended use(s). Then, working in conjunction with the primary intended users, evaluation questions are identified and appropriate methods are selected and developed. All subsequent evaluation activities are focused on these primary intended users and uses, including data gathering, analysis, and presentation. Thus, the primary users are engaged in interpreting data and in making decisions based on data. Broader dissemination, beyond the primary users and uses, is considered after this primary use, since findings may be useful to others in similar situations.

For some UFE, the primary intended user and use may be to provide a funder with a summative judgment of the project. In this case, methods are selected to meet that purpose. This work may include annual progress reports, but the focus is on summative judgment of a completed, static project. The project team may have little interaction or conversation with the evaluation team outside of receiving copies of the annual and final reports.

Here, we discuss UFE in cases where the primary users are the project team members, and the primary use is for growth and improvement. Our goal is not to provide a comprehensive how-to guide for evaluating faculty development initiatives but rather to share lessons learned from our extensive experience. Our team has worked on STEM higher education in mathematics, biology, chemistry, Earth sciences, and across disciplines, addressing questions about student outcomes, instructional change, instructor beliefs, student attitudes, hiring practices, departmental change, and equity and inclusion (see: E&ER, n.d.). We conduct our own grant-funded education research and also serve as internal or external evaluators for others' grant-funded projects. Much of our work has examined professional development in instruction and other areas of STEM careers, and how it relates to broader organizational or societal changes.

We share examples from multiple projects throughout the chapter and we conclude with a case study of our most complete example of UFE, from a series of projects that provided faculty development workshops on teaching. We use the term "project teams" to include both the project leaders, or principal investigators on the grant, and the facilitators who run the workshops. They are distinct stakeholder groups even though some individuals may fill both roles. In contrast to summative, funder-focused evaluation, this type of evaluation incorporates formative feedback early and continuously throughout an ongoing, dynamic project. Evaluators may still provide

summative results to grant funders, other faculty developers, and directly to STEM faculty, but these considerations are secondary to formative use by the project team.

**Evaluation Process**

In STEM faculty development, many evaluation relationships begin when a project team contacts an evaluator while preparing a grant proposal. During the initial conversations with project teams, we discuss how we work as evaluators and come to an agreement on the scope of the evaluation work (e.g., Owen, 2020). We incorporate the concepts of UFE into our proposal both through our description of planned evaluation activities and a budget that aligns with those activities. It is crucial to begin with a UFE framework at this stage so that the budget supports the significant work necessary to co-create evaluation questions, methods, and reports with the team. The budget should also support the work of securing IRB approval. Each relationship is different and is based upon the needs and intended uses of the primary users, but we lay out the general structure and process in the following sections, highlighting approaches we have found to be helpful in evaluating faculty development initiatives and in building fruitful relationships with their leaders.

***Identifying Research Questions and Methods***

After identifying a primary intended user, appropriate research or evaluation questions are formed and methods to investigate those questions are selected (e.g., Patton, 1990; LeCompte & Schensul, 1999). Since grant proposals require identifying research questions and detailed plans of the project activities, these conversations occur early in the proposal development phase. Here flexibility with methods is key. For faculty development, evaluation questions almost invariably investigate changes in participants' thoughts or beliefs—and later, actions—throughout the program. If existing literature establishes what kinds of changes may be expected, surveys are commonly used since it is possible to write items about expected outcomes. However, if the evaluation aims to explore topics that are less well established in research literature, in-depth interviews or focus groups may be more suitable to help explore these new topics. Questions about the program's features are well suited to observing program activities, carrying out focus groups with the facilitators, or analyzing program documents. Questions about how programs affect participants' classroom practices suggest that classroom observations could be needed. We have adapted to investigate all these types of questions for different projects by

learning new methods, partnering with colleagues outside of our team, and even hiring to bring in colleagues with particular skills we need. For UFE, the methods are selected to fit the question(s) of interest, not the other way around.

**Timelines.** A particular challenge in evaluating faculty development is the long time frame needed for faculty to learn new practices, then put them to use. It may take even longer for them to become skillful with new practices or to see their benefits or shortcomings. If we are attempting to investigate such changes, we also consider whether to conduct repeated measures over time, use a retrospective pre-survey along with a post-test, or measure gains with a single post-test. In our experience, survey instruments are often reliable with STEM faculty, who tend to be more self-critical and use the full range of a scale; we do not observe the ceiling or floor effects that we've seen with other audiences. It is also helpful to consider the larger context when planning measurement timelines. In our experience, survey response rates drop during times of high societal uncertainty and anxiety (e.g., election cycles).

**Instrument design.** When designing data collection instruments and protocols, it is critical to make it as easy as possible for the respondents to complete them with fidelity (e.g., Bradburn et al., 2004; Dillman et al., 2014). For example, to limit survey fatigue, we do not ask participants to rate an exhaustive list of all program features. Rather, depending on the needs of the UFE, we may ask participants to rate a short list of key features or to describe a single best, worst, or most influential aspect of a program. We follow other best practices such as avoiding double-barreled items, which include two questions in a single item. Questions like this can frustrate respondents and lead them to abandon surveys partway through. Before administering surveys broadly, we refine items by pilot-testing with small samples of similar audiences and using think-aloud interviews (Charters, 2003) to help identify unintended interpretations of survey items.

Classroom observation protocols present a different set of challenges, especially if FD workshop participants are geographically distributed. It is not feasible to travel and record them ourselves, so we ship video cameras to participants to record their classroom teaching. Participants set the camera in the back of their classroom once a week to passively record and then ship it back to us with the set of recordings safely stored on the memory card. To keep procedures simple and to ensure that teachers follow through with recording, we use cameras with physical on/off and record buttons rather than complex on-screen menus. We also

developed a series of short videos to help instructors learn to set up the tripod, operate the camera, and place the camera in the classroom for the best sound. Together, such tactics make it very easy for participants to collect high-quality video observations with only about 1 minute of setup and teardown time each time they record. So far, around 70 participants across multiple projects have successfully followed this recording process with very little data loss due to collection errors. Investing effort up front to find, develop, and implement appropriate observation methods enables us to confidently answer research questions within the project budget. As instructional methods shifted to online and hybrid methods in response to the COVID-19 pandemic, many schools have invested in additional video technologies. It may be possible to passively gather classroom observation data from these existing sources, though it may be difficult for outside evaluators to gain access to institutions' internal platforms. In all cases, evaluators need to carefully consider and address issues related to privacy and data fidelity. For example, the presence or absence of students on recordings has implications for both IRB protocols and the types of questions that can be answered about instructional methods.

**Analysis Methods.** Data analysis methods are often selected at the same time that data collection procedures are developed. Later, new questions may also arise that require new or different analysis methods, whether from the evaluation team, the project team, or through their collaborative work. On one project, as we approached summative reporting, we initiated bi-weekly 'data conversations' with project leaders to discuss evaluation findings. These conversations raised new questions that led us to develop different ways to analyze the existing data.

Being flexible so we can apply multiple methods or draw on multiple data sources often helps us gain a deeper understanding. In looking at participants' instructional change following workshops, we collect self-report survey data and triangulate these data with classroom observations, interviews, and listening in on participants' follow-up discussions over email listservs. Combining methods lets us better evaluate the effects that FD workshops may have on changing participants' instructional practices and can lead to new insights. For example, in evaluating a biology initiative that paired undergraduates with mentors and supported them through undergraduate research experiences, we used an established student survey instrument as an evaluation tool and paired it with mentor interviews. The interviews revealed deeper insights

about mentor motivation and led to richer insights and a standalone research publication (Hayward et al., 2017).

***Needs Assessments***

One way to help focus evaluation efforts is to use a needs assessment to identify the most critical areas of interest or concern (e.g., Gupta, 2011). In the simplest sense, a needs assessment involves an evaluator listening to project stakeholders and helping them identify their most pressing needs. This may involve talking with the potential participants of FD projects or with the project team members, such as PIs, program facilitators, or other staff. Thus, the evaluator can attend to big picture themes and concerns while team members' attention is often consumed with planning for their individual roles and responsibilities or handling logistics and the day-to-day tasks of implementing FD activities.

For example, on one project we attended meetings as FD project team members laid plans for training people to design and lead a particular type of local educational program on their own campuses. The planners emphasized community-building in their broader goals for the project, hoping to generate an enriching professional community among participants over their shared interest, but their planned workshop activities were all content-focused. They had not considered how to actively build community beyond that which might happen naturally as participants engaged in content sessions together. We wrote a memo reflecting what we heard in the discussions and provided framing questions to help the planners reflect on their workshop design choices using a community-building lens, such as:

- How will the various program components contribute to the development and solidification of the community?
- What norms, values and perspectives unite this community? How might these inform your program design? What role will this program play in creating shared community norms, values, and perspectives?

This memo helped the team let go of some training content in favor of more explicit community-building activities. It also guided us in constructing survey questions about community connections for the subsequent evaluation and helped diagnose some deeper needs of the community as we observed the first workshops.

In another example, a grassroots scientific network had new funding to expand its professional support offerings to network members. Early in the new project, we conducted a set

of focus groups with its members and reported their thoughts on the positive and negative aspects of the network's offerings to date, and their ideas and concerns for the new project. This early needs assessment offered multiple benefits: For project leaders, some of these early ideas were galvanizing, as they developed professional development programming to align with member-identified needs. For members, seeing their ideas reflected in network activities built confidence that the leaders were listening to them. Other issues raised in the member focus groups, such as a perceived lack of diversity among the network leadership, proved to be difficult and important areas that required ongoing organizational attention. And for the evaluators, conducting a simple, early needs assessment not only enabled us to identify some key organizational issues to track but also helped to begin building a trusting relationship with the leadership team. We offered something immediately useful to them and signaled early on that even critical feedback could be offered in a supportive manner.

### *Building Relationships to Support Useful Evaluation*

This last example highlights that strong relationships are key to successful evaluation. Building rapport with participants helps improve response rates and make participants comfortable providing honest feedback. Results are then more representative and more complete. Building trust with project leaders helps create productive relationships where critical feedback fosters improvement and growth, not defensiveness. In real life, relationships with project leaders precede relationships with participants. But here we discuss participant relationships first and then move to project leader relationships since they are often more intricate and longer-lasting.

**Establishing Rapport with Participants.** Throughout faculty development projects, we take an active role as participant-observers, attending FD activities and building relationships with participants from the beginning. We ask project leaders to introduce us as evaluation partners during the workshop registration process. Leaders explain that our purpose is to evaluate the project, not the participants themselves, and clarify participants' role by providing a short overview of both content and timing. As a result, our study invitations are more likely to be read and responded to. We make the benefits of responding directly clear to participants: we provide pre-survey results to the leaders so they can plan workshops to effectively meet participant needs. For example, we may ask participants in an instruction-focused program to choose a

target course so that instructors can bring relevant textbooks or course notes for participants to use. These methods work: we consistently have 90%+ response rates on our pre-workshop surveys!

When possible, we attend workshops in person. In this way, we can observe and provide day-to-day feedback to facilitators. We also get to know the participants so that we are not just some disconnected evaluator sending survey invitations. Conducting end-of-workshop surveys in person helps ensure higher response rates than following up once participants have already returned home. When we participate in real time, we can converse with participants to explain our role as evaluators, discuss how their feedback will be used, and detail what we do to ensure confidentiality and anonymity before sharing the results. Indeed, participants have commented that this helps them see the value of providing thoughtful, candid feedback. The project team's attention can easily become dominated by those who offer feedback to them most vocally during events, but the evaluators' more neutral position affords more opportunity to seek varied feedback from a range of participants. For virtual workshops or in-person events we cannot attend, we conduct a video chat with participants at the end to discuss our role and distribute surveys. Like pre-surveys, we routinely reach 90%+ response rates on end-of-workshop surveys. We believe these relationships contribute to high response rates on later activities too, such as interviews or follow-up surveys. For follow-up surveys, we normally achieve response rates around 70% or higher.

**Creating Productive Relationships with Facilitators.** We start building relationships with facilitators early on, ideally, during the planning stages. This way, we can help the teams incorporate research-based strategies into their plans and we can also co-develop our evaluation approach and instruments so that the data we collect will be useful. We may attend planning meetings to share features of successful FD based on our prior evaluation work and discuss with facilitators how they will implement these features in their own plans. In the past, we have attempted to do this with one-page memos shared with project teams during planning, but found that they did not have much of an impact unless used to guide a conversation.

During in-person FD offerings, we conduct quick formative evaluations to help with day-to-day feedback and planning. One favorite is "Gots & Needs" where participants use sticky notes to anonymously share one thing they "got" from the day and one thing they still "need" or a question they still have. After the day ends for participants, we sit down with the facilitators to

quickly read and sort the comments to identify common themes. This activity provides facilitators with positive feedback (the Gots) and targets topics to cover during the next day or future improvements to make (the Needs). Participant needs can therefore be addressed quickly, making facilitators' responsiveness visible and tangible to participants. This is one of the quickest and easiest ways to help facilitate data-driven decision making in FD projects.

In multi-year projects, it is empowering to track when faculty developers use evaluation evidence to make changes and to measure the subsequent impact on program outcomes. In the best-case scenario, we can detect the footprint of these changes and help faculty developers iterate over time to make continued improvement. Strong relationships are crucial for us to gain a deep understanding of facilitators' growth and development; when they feel comfortable with us, they are able to share openly about their successes and their struggles as they work to implement and improve program offerings. When we share feedback, we highlight successes and strengths seen in the evaluation evidence, to ensure that effective features of the FD are continued. This also helps us to more clearly prioritize issues of concern seen in the evidence since the balance helps to temper critical feedback and place it in a larger context. One facilitator likened it to reading student evaluations, explaining that,

> *I think one of the things that was particularly helpful about [the evaluation] is that it usually comes with a framing of, 'Here's my summary of what I'm seeing in this feedback.' Because it's like reading your student evaluation forms where you see all the bad stuff that they hated— but having somebody else read it and say, 'No, here are the big lessons that I'm seeing,' it really helps.*

In addition to balancing feedback, we also find it helpful to open conversations with questions rather than statements or data. Project teams are often aware of areas that need improvement, so inviting them to discuss these areas through questions, rather than assertions, results in less defensiveness and more willingness to troubleshoot actionable responses to critical feedback. For example, rather than diving into our interpretations of the findings first, we open discussions with questions like, "What jumped out at you as you read through the report?" With more critical findings, we may ask teams, "What do you think it was about this session that made participants feel this way?" These questions help facilitators to take the participants' perspective, and to see opportunities for improvement rather than feeling criticized. Sometimes these conversations also reveal

new information about incidents or interactions that help us to reinterpret evaluation data or place it in context.

***Interpreting Results with Primary Users***

To maximize use of evaluation results, it is important to consider the format and the content of evaluation reports. It is also important to consider how to discuss evaluation reports with primary users. Depending on how established a project is, interpreting results with primary users may include co-developing new analysis methods and report formats or helping users to interpret findings in formats that we have developed previously (e.g., Torres et al., 2005). Either way, it takes creativity and conversation with the users to develop effective formats and approaches, but the time invested is amortized when strategies are reused in later years or other projects. Even in established, long-running projects, new questions may arise that necessitate new analysis or reporting methods.

**Reporting Formats.** Evaluation reports often serve different purposes. As evaluators, we may want to back up findings and recommendations with a thorough discussion of methods and a complete analysis of the data informing the conclusions. But after devoting much time and effort to prepare detailed reports, it can feel like they go out into a void – it is unclear who, if anybody, reads or uses the findings. Perhaps the user has difficulty sifting through a long, text-heavy report to locate the actionable findings, or has less interest in the methods than in the implications. Over time, we have developed a few strategies (outlined below) to help maximize the use of reports by presenting them in focused and easily digested formats.

***Executive Summaries.*** Some projects involve several components spanning multiple years. The findings from such projects are voluminous and can result in book-length reports. In this case, we use a "catalog" approach, building a thorough record of our methods and analyses that serves as a permanent record for us and as a source to write shorter reports and prepare presentations. Building the catalog first helps us to streamline reports because we know we can use it to supply any details that the user wants to know, rather than trying to anticipate those needs and risk overstuffing the report. Shorter Executive Summaries provide broad outlines of the projects and summaries of the most relevant findings. For example, our summative report on the Inquiry-Based Learning (IBL) Centers Project (Laursen et al., 2011) ran to 175 pages. The key findings were highlighted in a 13-page Executive Summary, then further distilled into a two-

page summary (Laursen, 2015). Each document meets a different intended use for a different intended audience. The two-page document has been shared with hundreds of FD participants to build participants' belief in IBL methods by summarizing the supporting research, while workshop participants have used the Executive Summary to build support for IBL methods with skeptical colleagues in their departments. The full report is mainly used by those most interested in the methods and detailed findings, including researchers, grant officers, and the program leaders. Here the primary users had a goal of providing evidence about inquiry-based teaching methods that they hoped would be persuasive to others, so crafting evaluation findings into publishable research papers was also an important reporting strategy for this project.

***Dashboard Reports.*** For some FD projects, we want evaluation reports to serve as a formative tool and also provide benchmarking as new facilitators learn to implement an established program model. For formative evaluation, timeliness is of the highest importance so that data-driven adjustments can be made mid-project. We need to be able to collect, analyze, and report data back very quickly to facilitators and program leaders while the events are still fresh in their minds and so that they can follow up with ongoing participant needs.

To meet this purpose, we developed what we call a "dashboard report" (e.g., Hayward & Laursen, 2013; Smith, 2013). Rather than relying on long passages of explanatory text, it is a highly visual report template filled with charts. The entire process can be quickly replicated for each iteration of the program: we create a new data "collector" to send the same electronic survey to each new cohort of participants, feed the results into the data table of the report template, and generate a new report in a matter of days. Snapshots of these components are provided in Figure 1. After automatically generating the report, all that is left to do is a bit of fine-tuning so that it is ready for sharing. We also write some short, explanatory text to help interpret the graphical results and place them in context by comparing them to prior years. For example, it may not mean much to see a pre-workshop motivation rating of 3.81 climb only to 3.95 following a workshop. So we might explain that, "Motivation was already high prior to the workshop, which is expected given that participants have committed a week to learn to implement new instructional methods. This pattern is normal year after year and is not cause for concern."

**Commented [SKS1]:** Embed figures 1. a,b,c near this paragraph

While we generate such reports through Qualtrics survey software and an Excel template, similar processes could be developed through other software. Some statistical packages such as

SPSS and R even allow the user to save and rerun analysis scripts. Thus a new dataset can be analyzed quickly by simply rerunning all of the existing analysis scripts. The results can then be fed into a standardized report template. Reports in this standard format can be compared across time to help provide benchmarks and identify trends. We find that FD facilitators are more likely to read and discuss these short, visual reports compared to longer, text-heavy reports. Discussing and comparing reports over time makes improvements visible and helps identify opportunities for future growth. Therefore, the dashboard format helps us as evaluators to provide timely feedback and helps the users by making findings more actionable.

***Appendices.*** Standardized report formats can easily become repetitive. To remove repetition, we move some sections to a separate appendix that applies across the project. For example, for FD programs that repeat annually, it is not necessary to include a full description of the project and evaluation methods in every report; rather, they can be presented one time in an appendix that is cited in each report (e.g. Laursen et al., 2021). The user can focus on the new, individualized results for that particular report.

***Annual Summaries.*** When engaging in continuous improvement throughout a project, methods and processes are not completely repetitive, so we supplement the Appendix with an Annual Summary document. Individual reports offer results from each iteration, but the annual summary identifies patterns year-to-year (e.g. Laursen et al., 2021). We discuss how leaders used prior data to implement changes, what the new data indicate about those changes, and what recommendations we have for the next iteration. Again, the documents serve different purposes for different users. Individual reports are used by facilitators and program leaders to help facilitate the cycle of continuous improvement, while the annual summary documents the process, makes growth visible to facilitators and leaders, provides process-based lessons for other projects, and serves as a summative evaluation document for funders and other stakeholders.

***Other Formats.*** As projects continue, evaluation results may provide interesting new insights that are applicable for different audiences beyond the primary intended user(s). These can be shared in creative ways so that they are easily accessible. Products such as short videos (e.g., Yoshinobu & Laursen, 2021) or blog posts (e.g., Yoshinobu, 2021) can provide key findings more broadly and more quickly than traditional academic publishing can.

Taken together, these varied strategies provide a flexible set of choices that help ensure that relevant results are shared quickly and effectively with various audiences in ways that are appropriate and actionable. Recently, we've begun to supplement these passive reporting strategies with more active "data conversations" to help users interpret evaluation results and use data to plan next steps.

**Data Conversations.** Patton argues that one of the outcomes of UFE is that "as a result of being involved in an evaluation, primary intended users learn to reason like an evaluator and … [this increases] participants' capacity to use evaluative logic and reasoning" (p. 229). In an effort to achieve this, we try to involve facilitators in interpreting evaluation results and turning them into actionable next steps by engaging them in discussing the report. When we do, we give them time to individually read and process the reports before hosting a virtual meeting to discuss findings and plan changes together. During these meetings, we help the project team to make data-informed decisions and we provide insight from our broader experiences. By engaging directly with facilitators, we provide an opportunity to build a sense of ownership not just of their efforts on the FD itself, but also of the evidence and outcomes. Project team members may share evaluation results as part of their own professional activities such as blog posts, conference presentations, or events at their institutions.

**Conclusion: A case study of UFE for faculty development workshops**

Utilization-focused evaluation provides a framework to design and conduct evaluation work. Not every project evaluation will reach the same level of utilization, but when done well, UFE not only evaluates, but also helps to improve projects. We have presented strategies separately, so we close this chapter by sharing a successful example of how the strategies were used in concert to empower a team's continuous improvement efforts over the course of a project.

The example project was the third in a series of projects our team evaluated that used faculty development workshops to teach mathematics instructors to implement inquiry-based learning. The first project in the series offered four independently-developed workshops. The second project built upon the first by synthesizing best practices from the four workshops and from research to develop a single workshop model that could be replicated (Yoshinobu et al., in press). The third project focused on building capacity to deliver the workshops more broadly.

More than 25 faculty developers learned to implement the workshop model and then offered 14 intensive workshops over five years. These developers also collectively developed ways to reach new audiences through offering online versions of the intensive workshops and by traveling around the US to lead two dozen shorter workshops for instructors who weren't yet ready to commit to an intensive workshop but were interested in learning more.

Since our team had experience evaluating the prior IBL workshop projects, our first interaction with the project teams utilized the earlier evaluation results to help the teams learn about the workshop model and the theory behind its design. The model was fairly well defined with 4 "strands" serving various purpose, but teams had flexibility to customize it a bit by varying the sequence or content within each strand. (For a full discussion of the model, see Yoshinobu et al., in press). Sharing our broader experience across multiple projects helped teams to make modifications that still served the larger purpose of each workshop strand. For example, in some early workshops, participants told us on post-workshop surveys that they did not see the point of the "Reading Sessions" strand that shared research evidence in support of inquiry-based learning. Personally, they were already convinced by the evidence in support of IBL and had committed a week to attend an intensive workshop about it.

This finding might have encouraged facilitators to eliminate these sessions. However, data also showed that the utility of the sessions often became evident later, as participants reported using material from the Reading Sessions to help make the case for inquiry learning to skeptical colleagues or students back at their home institutions. Seeing the larger context over time helped the project teams decide to refine those sessions rather than get rid of them. They reframed the strand as "Literature to Practice" and broadened its original purpose, using research literature to not only build support for IBL but more deeply inform participant understanding of student learning. After the redesign, the post-workshop feedback on this strand was much improved and participants still found the evidence base useful in the longer term as well. Thus, the new teams not only learned to implement the established workshop model, but also improved it. By situating findings within the larger context, we helped the teams to preserve essential features of the workshop that could have been lost if changes were made based on only a single evaluation finding.

Since the project offered the workshop 14 different times, we used a dashboard report template to provide easily digestible results quickly (e.g., Archie et al., 2020; Archie, Hayward,

Daly, & Laursen, 2021). After we prepared each workshop report, we sent it to the lead team and the facilitators to review individually. We then hosted group sessions to discuss the report and make plans for improvements for the following year. We opened with questions rather than statements: we asked the facilitators for their thoughts and reflections on the report and the workshop itself. While debriefing evaluation results certainly has the potential to be difficult and emotionally charged, we found the opposite to be true. In fact, the meetings often included a good amount of socializing, so we started referring to these meetings as "data happy hours."

Each year, the teams used evaluation results to make some tweaks to their workshop implementation. We summarized the themes from the year and described these planned modifications in a document we called the "Synthesis and Study Methods" (Laursen et al., 2021). We added to it annually, updating with new themes and planned modifications and reflecting on the impact, if any, of the modifications from the previous year. This document catalogs the continuous improvement efforts throughout the project, including topics like reworking the Reading Sessions into the Literature to Practice strand, making deliberate efforts to provide synthesis and coherence across the strands at the workshop, and ongoing efforts to improve diversity and inclusion across the project team, the participants, and the examples of classrooms shared at the workshops. In capturing our understanding of the data each year, the document enriches and complements the final quantitative report on the project's achievements (Archie, Hayward, & Laursen, 2021).

Facilitators engaged in this 'cycle of inquiry' throughout the project, using data to evaluate and fine-tune their practices year-to-year. While the evaluation team captured these outcomes in our annual summary documents, over time, the project team expressed a desire to have greater ownership of the evaluation results and outcomes of the project. We hosted collaborative sessions to develop a handbook of best practices for running FD workshops. The handbook is filled with contributions from the facilitators based on their experiences and supported by data collected by the evaluation team (E&ER & AIBL, 2021).

Facilitators also took ownership of some of the formative evaluation tasks as the project progressed. For example, in the beginning, an evaluation team member conducted the "Gots and Needs" activity with participants at each workshop as a daily formative assessment. In later years, facilitators took on the "Gots and Needs" themselves. In the beginning of the project, each facilitator was responsible for running a single strand. Someone from our evaluation team

observed and helped the facilitators to notice things that sometimes could get overlooked as they focused on the logistics of running their separate strands: quiet participants, themes that recurred across strands, or topics that needed to be revisited in more depth. As the project progressed and teams grew, facilitators also took on this role of observing and synthesizing. The facilitator assigned to this role observed throughout the day, offered big-picture insights to their fellow facilitators, and provided synthesis for participants via morning introductions and afternoon debriefs. Sometimes more experienced facilitators would fill the role to let others run strands and other times, new facilitators would fill this synthesizing role as a way to observe how the entire workshop functioned before they were responsible for running one of the strands in later workshops. In general, we find that as facilitators become more familiar with evaluation and begin to think like an evaluator, internalizing how the feedback cycle improves their work, they are able to take on some formative evaluation activities that use less formal data collection methods and enable rapid refinement.

The collaboration between the evaluators and project team catalyzed improvement in both directions. Conversations with the project team helped us to improve our methods to make the evaluation more useful. The workshop model includes ongoing follow-up activities conducted via a group email listserv in the year following each workshop. Originally, we used frequency-based metrics such as number of messages per participant. But these did not answer the question the team was most interested in: what practices can facilitators use to create productive follow-up conversations? To answer this question, our group developed a new analysis method to code not just the frequency and volume of interaction, but the content of those interactions. Our team had to develop new skills to analyze relational data via social network analysis methods (Hayward & Laursen, 2018). We used the lessons learned from this analysis and shared it with the facilitator teams at an annual planning meeting. The facilitators then used these findings to develop strategies for more productive listservs. We found, for instance, that conversation prompts timed to align with participants' contexts were more likely to generate responses and create a discussion than did more generic prompts. Timely prompts include topics such as asking participants about their first day plans at the start of the academic year, or about strategies for gaining student feedback on their course towards the middle and end of the term. At the planning meeting, the teams incorporated this finding by creating a calendar of timely conversation starters to use throughout the upcoming year.

Over the course of the project, we learned a lot about successful faculty development and wanted to collaborate with others involved in similar work. So as the end of the project approached, the project team and evaluators co-hosted a summit for discipline-based professional developers across a variety of STEM disciplines. About 50 faculty developers met for 3 days in fall 2019 in order to exchange strategies and build collaborative relationships (AIBL, 2019). Though this one-off event required significant grant support, similar events may be useful every few years as the field continues to develop and grow.

On the whole, implementing UFE on this project supported the project team's efforts to grow and improve. They used data both to plan changes and then to reflect on the effects of those changes in subsequent iterations. A facilitator on a workshop project explained during a focus group that,

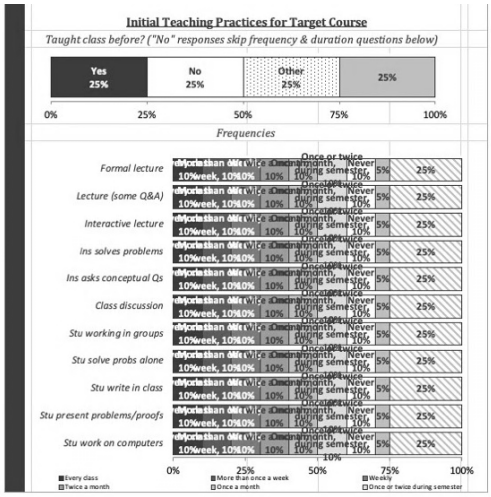
> *As someone who is not on the lead team for this grant, it has been incredibly helpful to me to see the evaluation side. It makes me think that as a principal investigator, perhaps I should be sharing my evaluation pieces with my [team]…. [The evaluation team's] communication piece is really important for people who want to improve in this work. It's important to me that your communication doesn't just go to project leaders but it goes to people who are involved at sort of the next level down in terms of [leading the FD].*

UFE has shifted this facilitator's view: They no longer see evaluation as a summative judgement for project leaders or funders; instead, they see it as a tool for improvement to be leveraged by the entire project team. They have learned to "reason like an evaluator."

**Figures**

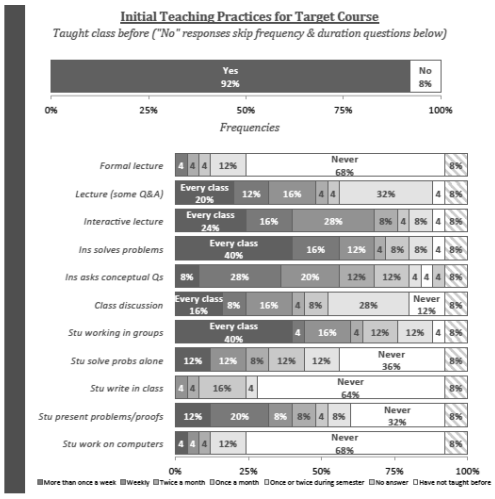
FIGURE 1. Dashboard template example.

(a)




(b)

| | Every class | More than once a week | Weekly | Twice a month | Once a month | Once or twice during semester | Never | No answer | Have not taught before | Totals |
|---|---|---|---|---|---|---|---|---|---|---|
| Stu work on computers | | | | | | | | | 25% | 25.0% |
| Stu present problems/proofs | | | | | | | | | 25% | 25.0% |
| Stu write in class | | | | | | | | | 25% | 25.0% |
| Stu solve probs alone | | | | | | | | | 25% | 25.0% |
| Stu working in groups | | | | | | | | | 25% | 25.0% |
| Class discussion | | | | | | | | | 25% | 25.0% |
| Ins asks conceptual Qs | | | | | | | | | 25% | 25.0% |
| Ins solves problems | | | | | | | | | 25% | 25.0% |
| Interactive lecture | | | | | | | | | 25% | 25.0% |
| Lecture (some Q&A) | | | | | | | | | 25% | 25.0% |
| Formal lecture | | | | | | | | | 25% | 25.0% |

(c) 


*This visual conveys participants' pre-workshop teaching practices much more concisely than text would. The blank report template of panel A is overly busy because it contains all possible labels. The evaluator copy/pastes the results from the data analysis software into the blank cells of the data table in panel B. This automatically creates the new visual that only needs a bit of cleaning up to remove extraneous labels before it is publication ready (panel C) and can be used to quickly identify patterns.*

**References Cited**


Academy of Inquiry-Based Learning. (2019). *Boulder Summit on Discipline-Based PD.* https://sites.google.com/colorado.edu/boulder-summit/home

Archie, T., Daly, D., & Laursen, S. (2020). *Collaborative research: PROfessional Development and Uptake through Collaborative Teams (PRODUCT) supporting Inquiry Based Learning in undergraduate mathematics. Combined evaluation reports: 2019 workshops.* Ethnography & Evaluation Research, University of Colorado Boulder. https://www.colorado.edu/eer/content/archie-combined-product-reports-2019

Archie, T., Hayward, C., Daly, D., & Laursen, S. (2021). *Collaborative research: PROfessional Development and Uptake through Collaborative Teams (PRODUCT) supporting Inquiry Based Learning in undergraduate mathematics. Follow-up report: 2019 workshops*. Ethnography & Evaluation Research, University of Colorado Boulder. https://www.colorado.edu/eer/content/archie-workshop-followup-report-2019

Archie, T., Hayward, C., & Laursen, S. (2021). *Collaborative research: PROfessional Development and Uptake through Collaborative Teams (PRODUCT) supporting Inquiry Based Learning in undergraduate mathematics. Cumulative evaluation report: 2016 – 2020.* Ethnography & Evaluation Research, University of Colorado Boulder. https://www.colorado.edu/eer/content/archie-cumulative-quant-report-product-workshops-july2021

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design--for market research, political polls, and social and health questionnaires.* John Wiley & Sons.

Charters, E. (2003). The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice, 12*(2). https://doi.org/10.26522/brocked.v12i2.38

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method.* John Wiley & Sons.

Ethnography & Evaluation Research. (no date). *Ethnography & Evaluation Research: Research areas.* https://www.colorado.edu/eer/research-areas

Ethnography & Evaluation Research and the Academy of Inquiry Based Learning. (2021). *AIBL handbook for online professional development: Lessons learned from PRODUCT workshops*. https://tinyurl.com/AIBLHandbook

Giersch, S., & McMartin, F. (2014). *Promising models and practices to support change in entrepreneurship education*. Epicenter Technical Brief 2. Stanford, CA, and Hadley, MA: National Center for Engineering Pathways to Innovation.

Gupta, K. (2011). *A practical guide to needs assessment.* John Wiley & Sons.

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching*, *8*(3/4), 381-391. https://doi.org/10.1080/135406002100000512

Guskey, T. R. (2000). *Evaluating professional development*. Corwin Press.

Hayward, C.N., & Laursen, S.L. (2013). *Collaborative research: Supporting Pedagogical Innovation for a Generation of Transformation via inquiry-based learning in mathematics (SPIGOT). Evaluation report: Workshop 1 at California Polytechnic State University, San Luis Obispo, June 24-27, 2013*. Ethnography & Evaluation Research, University of Colorado Boulder. https://www.colorado.edu/eer/content/hayward-calpoly-ibl-workshop-report-2013

Hayward, C.N., & Laursen, S.L. (2018). Supporting instructional change in mathematics: Using social network analysis to understand online support processes following professional development workshops. *International Journal of STEM Education, 5*(1), 1-19. https://doi.org/10.1186/s40594-018-0120-9

Hayward, C.N., Laursen, S.L., & Thiry, H. (2017). Why work with undergraduate researchers? Differences in research advisors' motivations and outcomes by career stage. *CBE—Life Sciences Education*, *16*(1), ar13.https://doi.org/10.1187/cbe.16-07-0229

Kennedy, M. M. (2016). How does professional development improve teaching?. *Review of Educational Research*, *86*(4), 945-980. https://doi.org/10.3102/0034654315626800

Laursen, S. (2015). *From innovation to implementation: Multi-institution pedagogical reform in undergraduate mathematics. A summary of research results.* University of Colorado Boulder, Ethnography & Evaluation Research. https://www.colorado.edu/eer/node/350/attachment

Laursen, S., Hassi, M.L., Kogan, M., Hunter, A.B., & Weston, T. (2011). *Evaluation of the IBL mathematics project: Student and instructor outcomes of Inquiry-Based Learning in college mathematics.* University of Colorado Boulder, Ethnography & Evaluation Research, Assessment & Evaluation Center for Inquiry-Based Learning in Mathematics. Full report: https://www.colorado.edu/eer/content/ibl-math-report-all050211 Executive summary: https://www.colorado.edu/eer/content/ibl-math-exec-summ-050511

Laursen, S., Archie, T., Daly, D., & Hayward, C. (2021). *Collaborative research: PROfessional Development and Uptake through Collaborative Teams (PRODUCT) supporting Inquiry Based Learning in undergraduate mathematics. Synthesis and study methods: Final version*. Ethnography & Evaluation Research, University of Colorado Boulder. https://www.colorado.edu/eer/content/product-final-synthesis-report-aug2021

LeCompte, M. D., & Schensul, J. J. (1999). *Designing and conducting ethnographic research (Vol. 1)*. Rowman Altamira.

Owen, J. M. (2020). *Program evaluation: Forms and approaches*. Routledge

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, inc.

Patton, M. Q. (2003) Utilization-focused evaluation. In T. Kellaghan & D .L. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 223-244). Kluwer International Handbooks of Education, vol 9. Springer, Dordrecht. https://doi.org/10.1007/978-94-010-0309-4_15

Smith, V. S. (2013). Data dashboard as evaluation and research communication tool. In T. Azzam & S. Evergreen (Eds.), *Data Vizualisation Part 2.* (pp. 21-45). New Directions for Evaluation, no. 140. Wiley Periodicals, Inc.

Torres, R. T., Preskill, H., & Piontek, M. E. (2005). *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. Sage.

Yoshinobu, S. T. (2021). *How much professional development is enough?* The IBL Blog. http://theiblblog.blogspot.com/2021/01/how-much-professional-development-is.html

Yoshinobu, S., & Laursen, S. (2021). *IBL workshops work!* 2021 STEM For All Video Showcase. https://stemforall2021.videohall.com/presentations/2153

Yoshinobu, S., Jones, M.G., Hayward, C.N., Schumacher, C., & Laursen, S.L. (in press). A broad doorway to the big tent: A four-strand model for discipline-based faculty development on inquiry-based learning. *PRIMUS.*