EVALUATION ROUNDTABLE

When opportunity knocks, open the door: Evaluation amidst transition at the Colorado Health Foundation

Susan Parker

June 2019

Introduction

In the ten years from 2008 to 2018, the Colorado Health Foundation underwent an extraordinary amount of change in nearly every possible way.

The foundation's grantmaking grew from $26 million annually to $86 million. It changed its operating status, undertook two major strategic overhauls, grappled with a damning staff survey on an unhealthy work place, and saw its long time CEO leave and one with an entirely new vision take the helm, upending expectations and requiring new skills of all staff.

In the midst of this, the foundation created a research and evaluation department from scratch. While building out this brand-new department, its directors and staff had to continually re-shape their role in response to constant upheavals. They had to gauge how much to push and assert themselves, and when to back off. Evaluation directors and staff brought different views about their department's role and approach, and were able to exert varying degrees of influence not only on the foundation's approach to evaluation, but also its approach to strategy.

Historically, evaluation has presumed a certain amount of stability in the work. The field treats significant transitions in the organizational environment in which evaluation occurs as rare disruptions that evaluation directors and external evaluators have to scramble to respond to or wait until the dust settles. Yet Evaluation Roundtable benchmarking data show that transitions are more the norm rather than the exception in foundations. New CEOs come, strategies are revised, staff are restructured, roles are re-defined, and new values frames are adopted at a surprising frequency and pace, all ostensibly to improve the foundation's ability to achieve its vision of change.

Amidst such transition, evaluation can quickly lose its relevance. Evaluation staff and consultants often struggle to gain the attention of change-weary—or even the most change-energized—strategists. Given that reality, how can the evaluation field more effectively prepare for and manage these transitions while maintaining the quality, rigor, and value of evaluative thinking and data?

This teaching case explores transitions on a number of levels—strategy redesign, culture and values shifts, new expectations and assumption about roles and performance, and a new CEO and direction. It invites reflection on three dilemmas that mark times of significant organizational change and that are exacerbated by the dynamics of uncertainty, conflicting messages and mindsets, and unclear performance expectations:

- What is the appropriate role for evaluation and evaluation staff in shaping a foundation's emerging approach to strategy, particularly in relation to the role of program staff? To whom and what are they accountable?

- How can evaluation leaders navigate the competing interests, overlapping mental models about evaluation, and change fatigue that characterizes times of transition, so that they can stake out a high-value role for the evaluation function?

- Under what conditions are external evaluation consultants able to provide real strategic value to foundations while so much churn and uncertainty is happening behind the scenes, often unbeknownst to consultants?

This case is really about how evaluation leaders—working within or with foundations--can play a role in managing transitions *not just defensively, but proactively and productively*.

The Foundation's Early Grantmaking and Evaluation Mindset

In 2008, the Colorado Health Foundation was at the beginning of an ambitious new vision of making Colorado the healthiest state in the nation. Its CEO Anne Warhover, who had a background in commercial banking, had been preparing a shift in the thirteen-year-old foundation's operating status from a public charity to a private foundation, increasing its annual grantmaking from about $3 million to $42 million, and organizing its work into three programmatic areas: Health Care, Health Coverage, and Healthy Living. The foundation's grantmaking was mostly responsive and open to any 501(c3) organization working in Colorado in its three program areas.

The foundation's vice president for operations, Shepard Nevel—an attorney with a background in policy and business, but not in evaluation—had been tasked with the creation of the foundation's first evaluation approach. Nevel and Warhover were both interested in the concept of a set of simple metrics to track performance across portfolios of grants, not unlike the performance approaches common to the private sector from which they both came.

The approach Nevel developed for the foundation, which he called "measurable results," was built on the idea that a small set of quantitative measures could be used across all of the foundation's grantmaking to summarize impact for each program area, and could, in theory, be mined for information to inform strategic decisions. All grantees (except for those few that did not provide direct services) reported against one or more of these indicators, which were then "rolled up" by foundation staff to represent the aggregate impact of the foundation's grantmaking. Foundation staff settled on 12 measurable results against which to assess foundation-wide performance.

Foundation leaders were proud of their approach, promoting it as an innovative measurement model. It was featured in convenings and publications for both Grantmakers for Effective Organizations (GEO) and the Center for Effective Philanthropy (CEP). They had invested not only the organization's resources in this approach, but also its public profile.

In a 2012 blog post for CEP[1], Warhover wrote, "Instead of relying on extensive grant-by-grant specific metrics, we have developed and continue to refine an evaluation model that emphasizes clarity and simplicity – with the intention of reducing the data-collection burden for grantees. We have done so by focusing on collecting fewer, more meaningful data points. Along with providing guidance for our grantees, these results provide us with valuable lessons on the most effective use of our funds. Each Measurable Result is research-based and highly predictive of a healthy outcome.

[1] https://cep.org/how-evaluation-measures-up-a-ceos-perspective/

For example, research shows that increasing physical activity (one of our Measurable Results) is associated with managing a healthy weight. Thus, numeric increases in this Measurable Result help us track our progress in that area of focus."

Nevel hired an evaluation director—Alexis Weightman, who would later move into the policy department after two years—to implement the approach. She was followed by Marisa Allen. Both had worked as professional evaluators.

After taking the helm of the evaluation department, Allen quickly noticed that the measurable results the foundation had selected were, in fact, largely outputs. Allen found herself unable to answer the questions she and the evaluation team were being asked by the board and program staff. The data were not able to answer questions about 'why' or 'how,' making it less relevant for decision making. As Allen recalls, "the demands from staff were: 'I really need to understand how to change the strategy next time around [and] what are best practices that grantees are using?' while the question the output model was answering was 'what is the reach of the grantees?'"

Weighing how much change to push for

Allen happened to start her job at the foundation in the midst of a big change that exacerbated the pressure to answer questions that the measurable results could not. In 2011, foundation leadership sold its interest in the HCA-HealthONE joint venture and ultimately converted to private foundation status in 2016. That sale dramatically increased the foundation's assets—from just over $774 million to $2.2 billion. The sale meant that the foundation would soon have a much bigger profile in Colorado, bringing more scrutiny and a fair bit of anxiety and uncertainty to the staff, who were not quite sure what this all meant for their roles.

As she thought about how to shape the evaluation function at a foundation that would need to do significantly more grantmaking, Allen weighed a number of considerations. Those included that the foundation leadership, including her boss, often promoted the benefits of the current model, and the foundation was in a state of tumultuous change that would require rapid growth in staff. She chose not to jump into creating a more robust outcome framework and evaluation approach, but rather to slowly educate staff and the board about what evaluation could and could not do, in part to expand thinking about evaluation beyond the measurable results approach. She also shepherded the foundation's first external evaluation to demonstrate what was possible to learn through evaluation.

"It could have been a function of my style or a function of what I felt like would be successful, [but] I asked myself is there a way to keep measurable results and add something of value as opposed to changing something altogether and throwing things out?" Allen said. "The other thing that influenced my thinking was there was a lot of discussion about rapid change and how the staff had large workloads. And I noticed that we were not doing change management well. I think I thought, gosh, if I as the evaluation director add another thing that might be a bad decision.…What I heard was that more change could break the backs of program officers. I'm not sure how much that influenced me, but I was definitely aware of that."

Growing the department and making a key hire

In response to a growing organization, Allen hired two more people for her department including Kelci Price, whom she brought in as a senior evaluation officer in July 2012. In early 2013, Allen announced that she was leaving and Price was named as director of the research and evaluation department.

When Price first came on board, she was struck by the culture of the foundation.

“There were a lot of dry runs and prep sessions for going to the board and the philanthropy department (the name for the department overseeing grantmaking),” she said. “We had a culture that was very much about saying things in what I would characterize as a perfect corporate way. You would get feedback if there were mistakes in your write up or if you didn’t have answers to questions. It was really not accepted to say ‘I don’t know, that’s a great question.’ The staff needed to show up as experts. You had to have a very high level of certainty of what was going to happen because of this work.”

In this context, Price experienced the measurable results approach as an exercise through which executives and program staff could signal to the board that grantmaking decisions were rational and that they were paying attention to results. In reality, beyond program staff using the data to make adjustments to individual grants, few decisions were made based on measurable results data, despite how the approach was publicly cast.

Strategy Re-design

In 2012, leadership decided to undertake a strategy refresh. It had been more than five years since the first (and last) strategic plan, and the board and CEO were not sure the foundation was getting the outcomes they wanted.

Progress toward the foundation’s targets had proven slow and difficult to attribute to the foundation’s grantmaking. As many other large foundations had moved toward strategic philanthropy, the foundation’s executives and board were starting to question whether a responsive grantmaking approach was effective. In its smaller pockets of strategic work, the foundation had seen real evidence of change that could be connected back to grantmaking. For example, one of only two external evaluations the foundation had commissioned concluded that the foundation’s school-based health centers initiative had helped all of the funded centers become more financially sustainable.

A larger body of strategic work to increase health coverage and enrollment had seen gains too. But with responsive grantmaking, it was much harder to detect “impact,” let alone claim that it was due to the foundation’s support. Grantmaking felt too diffuse. And though the measurable results model had been able to give foundation executives a sense of the number of people grantees served, it did not provide evidence that the foundation was having the impact it wanted on the health of Coloradans. The executives and board chose to push for a more strategic approach to its grantmaking overall.

"The strategy refresh created the perfect playing ground for us"

The CEO gave Nevel's department–and ultimately the research and evaluation team—responsibility for ensuring each of the three program areas was on track with its own strategic planning process. Price, who had spent a good deal of her initial months at the foundation locked away in an office compiling grantee-reported data for the measurable results reports, had her own ideas about evolving the foundation's evaluation practice alongside the new programmatic strategy. For Price, the strategic refresh was an ideal opportunity to introduce new approaches to evaluation and learning.

"The strategy refresh created the perfect playing ground for us," she said. "We were all in flux. It was an opportunity for us to…do something totally different. Instead of having a steady state and trying to evolve evaluation in what would have felt like a piecemeal way, this created the context where nobody minded that we were changing…because everything else was changing too. It was an opportunity to create that big evolution without raising anybody's hackles."

Beyond the compilation of data for board reports, the research and evaluation department's stance had largely been reactive, waiting for people to have questions that the team could attempt to answer through some kind of data analysis. Price and her team envisioned taking a much more proactive role centered on the idea that learning is indispensable to strategy. They wanted to plant the idea that strategy was an experiment, and that the foundation needed to learn and improve it over time. They also wanted to increase the use of evidence in decision making. And finally, the team wanted to play a role in improving strategic thinking, in part by helping staff develop skills to ask more critical questions about their work. But moving from where they were now to such a robust evaluation and learning function would require proving that they could offer real value to the organization.[2]

Introducing theories of change

As an early step in expanding evaluation's role, Price and her department tried to support the strategy refresh by offering to facilitate theories of change for each program area, which the foundation have not previously used. In fact, Price said for years the foundation had not allowed the phrase to be uttered and remembers Allen warning her that the foundation was not on board with theories of change. Allen recalled several staff saying that they had seen many logic models or theories of change that were useless, with pages and pages of nonsense that were a complete waste of time.

The foundation had an aversion to "over-complicating" tasks and put an emphasis on doing things quickly, Allen recalled. These values and mental models overshadowed the need to look at the complex roots of social and health problems. The theory of change process was viewed as a burden rather than a tool for thinking clearly and logically about the work of the foundation.

To slip it in under the radar, Price named her first iteration of theories of change "outcomes frameworks," which was far more palatable to an executive team that was accustomed to using results and outcomes language.

[2] For an overview of key evolutions in the foundation and its evaluation and learning function, see appendix A.

With the impetus of the strategy refresh and the blessing of leadership, Price was excited about evaluation playing a role in strategy development. Program staff were immersed in gathering information about their areas of work, contracting for landscape scans, and reviewing existing literature in search of clues about what kinds of solutions might be most impactful. But how that study process was translating into a strategy was unclear. To Price, it seemed impossible to be thoughtful about strategy if foundation staff hadn't articulated their thinking. And better strategic thinking would lead to better impact. What's more, at the time there was essentially no documentation of the foundation's strategic approach and assumptions, which also made it difficult to structure evaluation and learning. Price and her team thought they could help solve these problems by bringing evaluative tools like theory of change to bear on the process. The evaluation team designed a theory of change process that they thought would help teams think big and broad about what it would take the make change happen, and then narrow down to the specific areas where the foundation could focus its work.

Because foundation leadership had not required a specific approach for the strategic refresh, instead leaving it up to individual portfolio directors to select the process they wanted, Price did not feel she could force the use of theories of change, particularly given the foundation's long-standing aversion to them. Instead, she and her team asked program teams to let them facilitate theories of change with the promise that it would help them with their strategy process by revealing gaps in their thinking. Price also emphasized that the evaluation team needed theories of change for evaluation purposes.

At first, the teams were excited, meeting in the early hours before work and plastering the walls with sticky notes. But then the ideas kept coming and the font on the theory of change documents got smaller and smaller while the number of pages increased. When one team's theory of change had to be put on a 6' x 4' poster with no indication of stopping, Price said she started to get very nervous. She worried that it was too detailed, too linear, and projected too much certainty. She was also concerned that the team was getting so caught up in the details of the work that they could no longer explain the overall vision for what was going to change and why, nor what the foundation's *specific* role might be in that change, which she saw as crucial for effective strategic planning.

Who and what is the theory of change for?

"One particular meeting is seared in my memory," Price recalled. "Since my major issue was getting the teams to take anything <u>out</u> of their theories of change, for this particular meeting I decided to have the team talk about just one strand. I was feeling good about this—I walked into the room and the entire team of 8 people was there. I stood at the front and said, 'Today we're just going to focus on this one strand and we're going to unpack that a little bit.'"

"And I smiled as I waited for their response. And things turned ugly. The team lead said to me: 'We don't want to do it this way, we don't think about it like that.' Another team member chimed in: 'You don't seem to understand, this is all connected!' And they glared at me. And I replied very intelligently: 'um uhhh, okay.' I was stunned. I wasn't prepared for this response and I didn't know what to do. I stuttered and stammered and stared at the clock. Still one hour and 57 minutes to go…. The next week I suggested to one of the team leads that I was going to do some work on the theory of change before our next meeting. She said: 'You know what, I think we'll edit our own theory of change files from here on out. You let us know if you want to see them.' So I backed off."

Even though Price's team was unhappy that the theories of change process did not sufficiently unpack staff's thinking about the foundation's specific role in change, they realized that the program teams were not ready to further refine them in the midst of a demanding strategic refresh.

Erica Snow, who is a portfolio director at the foundation, recalls the work on theories of change this way:

"My sense is that [senior program] staff really saw it as separate [from their own strategy development process]," said Snow. "Program staff never took it on to own that jointly. It was like, that's their [the evaluation department's] thing and they're going to measure us and they're going to come in at the end and sort of map out this pretty theory of change to tell us this is the outcome and here's how we're going to get there. I'm not sure if programs ever really saw it as…an important tool that you bring in early on as you begin to map out the strategy. It struck me that institutionally—across the organization--there wasn't a shared value around the purpose and use of the theory of change."

Program officer Hillary Fulton said that she had a different discomfort with the work around theories of change. Input from others outside the foundation with respect to what the foundation ought to focus on seemed limited, and she wondered whether foundation program staff really ought to be the ones developing the theories of change. She felt the theories of change turned into a long list of all the things it took to reach an outcome, but did not help staff know more about what they should *do* specifically to reach those outcomes. Without a more inclusive process, she felt less competent to be making choices about what the foundation's strategy should be.

For Price, the disconnect came because her team wanted to use theories of change as a way to pressure test, develop, and improve thinking *during* strategy refresh, while program staff seemed to value them only as visual representation of their strategy that they could give to executives *at the end* of the strategic refresh process.

What do we mean by strategy?

On top of that, Price and her team jumped into a process of articulating and unpacking program staff's strategic thinking when staff were still trying to figure out what it even meant to develop "strategy." Since the foundation primarily engaged in responsive grantmaking, there was no institutional practice around what strategy should look like. The evaluation department staff themselves were novices at using theories of change to develop strategy, having used them only on the project or program level, which felt more linear and clear. Articulating strategy, in Price's view, would require a different way of thinking that no one was entirely prepared or trained for.

"One of the things that I realized is there are a lot of things about evaluation that are not designed to support the kind of work that goes on in philanthropy," Price said. "Evaluation is set up to evaluate programs and it does that well. You get up to the strategy [level] and evaluation doesn't have great tools for that, particularly when you use a theory of change in which things are predicated [to happen] in a linear way. What makes a good strategy is not something they teach in an evaluation program, where it's all about methods. We realized we really didn't know anything about strategy, but we were trying to capture people's thinking about strategy in these very linear theories of change."

Price said that when she started working at the foundation she believed that when you plan a strategy you need to be certain about its outcome and make a compelling case that it is going to work. But as she studied the literature, she learned that approach does not take into account complexity theory or emergent or adaptive approaches. Price studied Henry Mintzberg's work that says you do not plan a strategy, you *learn* a strategy, setting out with initial ideas, observing what's possible and how the system reacts, keeping some of those initial ideas and abandoning others, while adding entirely unexpected ways of working along the way.

Is it about product or process?

While some program staff did find the work in theories of change helpful in understanding strategy, most did not. After that less-than-productive experiment, Price and team switched tactics from trying to co-design a theory of change to simply writing down what they heard during strategy refresh meetings and giving it to the program staff as a way of capturing their thinking. It was an experiment to see if that approach was less burdensome and more successful, and staff seemed to find it helpful, Price said.

Eventually, she and her team realized that they needed to be more agnostic about how best to articulate the thinking behind the strategy and move away from rigid boxes and linearity. The team tried a variety of approaches including visuals, narratives, reverse visioning, and others. They tried to adapt to various groups' style of thinking with the overall goal of getting the chain of thinking and connections between pieces to be clear.

Price had a second chance to try theories of change in 2015 after the foundation had lived with the new strategies for a while. This time, their appetite and energy for the process was much higher. Price speculated that this was because the evaluation team had been working all along on helping staff sharpen their thinking through theory of change questions, even though they were not calling it that or using a recognizable theory of change format. In short, rather than imagining that their charge was to implement a specific technical process or generate a specific product (i.e., theories of change for all strategies), the evaluation team had come to operationalize their role in strategy as building capacity for strategic and evaluative thinking.

Reflecting back on missteps in their theory of change efforts Price said, "my major mistake was vastly underestimating the enormity of the change process we were undergoing as an organization. We had decided we wanted to be a more strategic grantmaker, but after years of doing responsive grantmaking it turned out we couldn't just flip a switch on the way we were used to thinking. I didn't realize how difficult it would be for our teams to pick and choose what the foundation should focus on—and what should be left out… It was also that we hadn't yet built our capacity as an organization to have the conversations and make the decisions a theory of change process entails."

Introducing emergent learning

When Price assumed her role as evaluation director, she was also eager to expand the evaluation department's role to more explicitly support learning. She viewed a quality learning practice as the critical linkage between doing evaluation and improving the impact of the organization. Many program staff were signaling eagerness to learn, but no real "muscle" had been developed to do so organization wide. At the time, learning tended to be viewed as an individual-level effort to acquire knowledge and skills, but Price wanted to recast learning as "using what we know to inform what we do," with specific attention to using evidence.

Shortly after starting her role, she learned about emergent learning tables[3] and later met Blair Dimock, vice president of partnerships and knowledge at the Ontario Trillium Foundation, who told her that using emergent learning had been so successful that staff stopped him in the hallway and asked for more learning sessions. Price immediately took notice.

"I abandoned my normal obsessive researching and comparing of models and thought: 'this seems compelling, let's try it,'" Price noted.

Price framed emergent learning as another way to help with the strategy refresh rather than as a separate learning practice that would come *after* strategy hit the ground. The evaluation department had noticed that the organization's strategy development process did not include a strong practice of reflecting back on their previous work or on evidence from the field and then figuring out how to incorporate that information into future work. Emergent learning offered a way to help staff surface and synthesize such thinking and evidence, and build a thread from what the foundation already knew to what it would test in the future.

She also liked that the approach requires explicitly naming hypotheses about what actions will lead to what results, which then become testable. Too often, Price said, foundation staff still framed an idea about what types of strategies are the "right" ones as truth, holding on to it tightly, defending it, and expecting it to remain true no matter what changed in an initiative or environment. When something is framed as a hypothesis, it helps put people in the mindset of thinking that something is testable (and this might not hold true), and they may need to tweak their thinking.

Price felt that theories of change provided a way to articulate thinking about how the world works and how a certain set of interventions might make changes. Emergent learning was a way to pressure test this thinking against evidence and experience without getting so committed to a particular way forward that it would shut down people's ability to recognize disappointing results, learn, and then adapt.

Getting immediate take up with emergent learning

At the end of 2013, Price brought in the consulting group Fourth Quadrant Partners to lead a learning session to help with the strategy refresh. The evaluation team introduced emergent learning by telling program teams that they could help facilitate an already-scheduled strategy meeting using a new technique, and promising that they would get much more value out of that meeting than they typically did.

---

[3] Developed by Fourth Quadrant Partners, Emergent Learning is a disciplined team learning practice that pushes participants to apply observations about the past (evidence) to future decisions and actions. The practice includes a variety of tools, including Before and After Action Reviews and Emergent Learning Tables. Emergent Learning Tables take learners through a process of 1) observing what happened on the ground and identifying what factors drove results; 2) drawing insights about what has worked under what conditions in the past; 3) given these insights, generating new hypotheses about what will make them successful in the future; and 4) identifying specific upcoming opportunities to put these hypotheses into practice (learning into action).

She got immediate take-up from participating staff who seemed to love it. Within a few weeks, they asked the evaluation department to bring the consulting firm back to do half day sessions, which Price said was unheard of at the time. Getting even just a 60-minute meeting for the evaluation team was a miracle.

This reflective practice was so valuable to the foundation that they began to apply it to other pieces of their work. The evaluation team was able to begin taking a role in helping the foundation reflect on and learn about their internal practices, processes, and structures, rather than focusing learning on just the content of a strategy. In a paper about the experience, Price wrote:

> As the Foundation moved towards initiative-based grantmaking, there was a need to figure out what the planning phase for initiatives actually looked like, and how the cross-functional departments would be engaged (i.e., policy, communications, evaluation). This process was fairly underdeveloped in the beginning, and early issues arose about how different teams were being engaged and at what point in the process. We recognized that EL [Emergent Learning] could help the teams discuss these issues and help people think through how to approach the next set of initiative planning (which was imminent). We facilitated After Action Review[4] (AAR) processes with each of the teams, then synthesized those findings to share back with the teams and with management. The AAR process helped the teams make explicit, synthesize, and document the challenges they had encountered and possible solutions. This proved invaluable for helping the teams navigate turbulent waters before organizational processes had been solidified (and for informing the ultimate approach to the processes). We witnessed concrete learnings and changes that came from these conversations in how teams interacted with each other during initiative planning. For instance, one director took responsibility for overseeing the development of a guide to initiative development, which had been identified as a significant missing piece. Other actions weren't necessarily codified into the process, but hearing stories of how some team members had not been effectively engaged raised the awareness of program staff and led them to change the way they interacted with their planning teams.

"There's a lot of positive energy"

One of the early champions of this approach was Hillary Fulton, who used it to help her in crafting both an initiative, Healthy Places, and an external evaluation of that initiative. Fulton was the first person outside the evaluation team to use the word hypothesis, which for Price was a tangible signal that ideas they introduced could diffuse into the thinking of program staff very quickly. Healthy Places was the foundation's first big initiative on the built environment and increasing physical activity in community settings. One of the program's goals was to learn and share lessons through a series of grantee convenings. Fulton decided to use emergent learning techniques to design each successive convening in a way that would improve on the last, based on evidence that the convening had or had not achieved the desired ends.

"What I like about the emergent learning tables is that it helps you think about all of the different data that you have available to you," Fulton said. "I liked that you can highlight what insights you

[4] After Action Reviews, a practice that is part of the Emergent Learning "suite" of learning approaches, takes staff through a set of focused questions: 1) What were our intended results? 2) What were our actual results? 3) What caused those results? 4) What lessons should we take forward next time?

are having and there's the element of celebration. It encourages participation and gives you the opportunity to build, rather than rehash the same conversations over and over. We really just hadn't had the structure to apply what we were learning very well prior to that. And they are pretty fun exercises. They draw out new ideas and there's a lot of positive energy. I like the energy of working collaboratively with people and building on ideas together. We do a lot of writing on our own, we do a lot of thinking on our own. There's a lot of responsibility that we shoulder to try to strategize the next best thing to do. So it feels good to be more like we're all in it together."

For Erica Snow, the emergent learning sessions helped provide space for staff members to make their thinking visible and challenge their mental models. At the time, she and Amy Latham, now vice president of philanthropy, had been working in health coverage for years while another staff member, Kyle Sargent, a program officer, was newer to the team. No formal mechanism existed for knowledge management. The sessions provided her and Latham a space to download a lot of the know-how stored only in their heads. Snow also said that she came in with a strong viewpoint about what works in the area, but the sessions expanded her thinking and, by making everyone's thinking explicit and visible, revealed where she and her colleagues were not in alignment.

For Price, the emergent learning approach was the beginning of what she envisioned as a long-term learning practice at the foundation. Over time, she would need to figure out how to use these techniques to help the foundation make sense of the evidence it was gathering from evaluation and apply it to its strategy choices, rather than focusing only on improving shorter-term activities. But for now, it was a start.

Role confusion and change

For the program staff, the strategy refresh was triggering a major shift in roles and expectations. They had all been hired as responsive grantmakers whose primary responsibility was to vet individual grant proposals and decide if the project merited the foundation's funds. But as the foundation re-imagined what it takes to have significant impact on large-scale problems, the scope of responsibility for program staff was changing dramatically. They became strategy designers themselves, responsible for analyzing the conditions out in the world, deciding what solutions the foundation should work on and how, and then determining a set of actions to take that would result in greater impact than they had been seeing with their responsive approach. What they were now accountable for looked very different, and they did not receive much help in developing the skills to carry out this new work.

As the program staff were working to understand and shape their new role in "producing impact," it was likewise a time of uncertainty about the role of evaluation. And none were clear about the relationship between the two. Although Price was jumping in and experimenting with a number of approaches to help clarify and pressure test program staff's emerging ideas, evaluation's role in the strategy refresh process was never clearly defined. Allen, who was still at the foundation at the beginning of the refresh, said, "one of the things that was left undone [when I left] was what was the appropriate role for the evaluation team related to strategy development? Is it a leading role? Is it collaborative?"

Trying to clarify the evaluation department's role

Price felt a similar lack of clarity. Throughout the refresh, strategies were conceptualized as something that the philanthropy department was in charge of and that the other departments such as evaluation and communications would support. The explicit charge for the evaluation department had been only to "keep track" of what the three portfolios were doing (timelines, consultants, etc.). However, Price *wanted* the evaluation department's role to include helping teams make their thinking explicit about the outcomes that they were seeking, the rationale for activities they were proposing, and the evidence that a particular approach would get them to a successful outcome.

For some program staff, too, evaluation and research's role were confusing, Latham said. "At best, maybe there was just a lack of clarity, and at worst a skepticism and a 'why are you asking all these questions' type of thing between the program staff and the evaluation staff. And it felt [to some program staff] rather than as a 'critical friend' kind of relationship, more of a questioning relationship. I think for some people perhaps they had deeply ingrained ideas about what evaluation does and it was from this accountability frame, as opposed to a learning frame. They might experience a meeting that evaluation was in and say they don't really add a lot of value or they're criticizing. We also had a culture of not being able to have [these] conversations in the moment."

At the same time, during the strategy refresh, some program staff embraced the role Price's team was playing in helping them develop and evaluate their strategies. They began to see the power in that partnership in their work, and those staff became champions and examples to their fellow program officers for the role evaluation could play, Latham said.

As ill-defined as it was, this period of hazy boundaries between roles gave the team a "field of play" where they could show up in new ways, posing questions about strategy, offering tools like theories of changes, and asking what success would look like. From Price's perspective, by the time roles were codified, the evaluation department had proven its value in those ways. If roles had been more tightly defined from the beginning, or if Price had waited for active permission to expand the boundaries of the evaluation department's role, they might not have had the opportunity to experiment and build value.

Organizational upheaval in the midst of strategy refresh

In January 2014, as the organization and the evaluation department were trying to chart a new path for themselves, a Center for Effective Philanthropy (CEP) staff survey was conducted internally that roiled the organization.

The survey exposed fractures in the culture that the executive team was unaware of and that had never been openly discussed. A number of people reported the culture as stifling, toxic, rigidly hierarchical, and expressed concern that issues of diversity, equity, and inclusion were not being addressed.

"In the middle of strategy refresh when the organization is trying to figure out how we do the work, all of a sudden we have this watershed cultural moment," Price said. "It made it that much harder because those wounds were now on the table and out in the open. And some of those wounds were very, very deep. It opened our eyes. There were some of us, like myself, who did not recognize the depth of the pain that people were feeling, the depth of the toxicity of the culture of the organization. All of a sudden we're aware of these tensions about the organization feeling too hierarchical, too top down. It contributed to our fears of putting any standardization in place. [As a result of the survey], the organization started leaning very hard into giving people a lot of power, a lot of autonomy to do whatever they needed to do so they didn't feel it was a top-down hierarchical process."

Price also noted that organizationally, the foundation has always had a fairly strong resistance to standardization of processes and expectations across different program areas and functions. From the beginning of the strategy refresh, program staff could approach the work any way they wanted to. The survey results made resistance to mandated approaches even clearer. In response, Price and her department decided to invite everyone to approach theories of change in the way that fit their needs. In retrospect, that decision proved fortuitous.

Price had come in with the mindset that the way to create an effective organizational model or approach for strategic grantmaking or for evaluation and learning is to do some early thinking and experimentation and then, after observing what works, put in place standardized processes and practices organization-wide. She later realized that it would have backfired for the organization to do that in the midst of all the upheaval, frustration, and pain staff were experiencing. The only standardization that Price ending up pushing for and getting was a requirement for each program to use a strategy template, which included a measurement and learning plan that described what evaluation and learning activities would take place around that body of work over a period of time, and a theory of change.

Deep-seated fear of failure

Among the effects of the culture exposed by the survey was a deep-seated fear of failure, even though the organization publicly presented itself as embracing failure, which many organizations such as CEP and GEO were urging foundations to do at the time.

"The narrative of the organization, the way we talked about ourselves, was very much like we're learning, we're failing forward, we're adaptive, all those kinds of things," Price said. "But the survey results suggested that actually people thought there would be negative personal consequences if they failed. They didn't trust that the organization was a safe space to learn and fail. That created an atmosphere where people were unwilling to fail."

The results of the staff survey continued to reverberate throughout the organization. By the end of the year, the CEO and a number of vice presidents had resigned.

This moment in the organization's history, like a number of others, was experienced by program staff as one in a long string of intertwined events and changes that, in retrospect, are difficult to untangle and recall. Staff were juggling a number of responsibilities and transitions, and as a result, the role of the research and evaluation department at this time faded into the background and more immediate concerns took priority.

"During this period my boss left the foundation and then I was moved into an interim role [taking that job]," Latham said. "It was an intense period of change for me in terms of the scope of my responsibilities…I'm really thinking about my own role and my team from a very pragmatic perspective about how we keep the work going in a way that does not cause disruptions to external partners."

Learning takes a back seat

In the midst of this uncertainty, Price hired Yen Chau as a senior evaluation officer. Chau had worked as an evaluation consultant, as well as a research analyst and director for schools and related organizations. Chau has a similar mindset to Price when it comes to managing change.

"I saw all of these changes like the CEO and other staff leaving as an opportunity," Chau said. "It's just my nature. I thought, 'okay now there are opportunities for us to propose and insert some things that I know I could do well and that the foundation might need.'"

But as staff anticipated the arrival of a new CEO to replace Warhover, Price observed the psychological effects of the many cascading changes and uncertainties they had been experiencing for the past few years. And the changes had only just begun.

"The biggest challenge I was facing at the time was that we were, organizationally and for my team, operating in a total 'in between' space," Price said. "We knew that things would change with new leadership but not how. This created challenges for us in thinking about what to assess through evaluation because we didn't know much about what the future direction would be. This also created challenges around learning because people were fatigued by all the change, and the change they knew was coming down the pike, so they weren't very interested in learning. Learning felt like an activity you do when you have some sense of direction, not something that was done when the future was totally unknown. I tried very hard to push the idea that learning was a practice that could actually help us navigate the in-between space and make good decisions about how to operate under conditions of ambiguity, but no one was biting. They were tired, they were not motivated to engage in that way, so we [the evaluation department] backed off and waited for the moment when learning would be of interest again."

A Broader Role for Evaluation During Strategy Implementation

The strategy refresh process ended in 2014 and the foundation moved on to implementation with a major organizational shift to more strategic grantmaking. This included an approach that was already fairly common in the field at the time, especially among larger foundations that viewed themselves as strategic grantmakers. "Funding opportunities" with requests for proposals had a structured set of funding criteria with more clarity about what the foundation was seeking to achieve and *how* they believed change could happen.

In this approach, nonprofit grant applicants would make the case for how their organization could contribute not only to the set of top-level results the foundation was interested in, but also how their strategic approach fit into the foundation's view of how larger-scale impact could be achieved. Program staff would assemble portfolios of grants that combine, for example, program delivery with communications and policy advocacy grants under the rationale that this combination of grants would presumably result in larger-scale impact than grants that were considered in isolation of one another.

In this approach, the program team's "value-add" in helping the foundation achieve its change goals was based not only on their ability to discern the quality of individual grant proposals, but more importantly on their ability to assemble a smart mix of strategies and grantees in a portfolio that would add up to more than the sum of its parts. The foundation also began experimenting with different types of impact approaches like program-related investments and collective impact. However, Price said that many program staff wanted to continue to fund organizations that they saw as worthy even if they were no longer well-aligned strategically. They also wanted to ensure that funding was still as widely available as possible so that potential grantees wouldn't be excluded as the foundation narrowed its criteria for what types of projects or outcomes it was looking for. As a result, they crafted some funding opportunities that were essentially still responsive.

A new vision for the evaluation function

The foundation's new vision for how it would engineer larger-scale impact opened the door for a new vision for the evaluation function as well. Price saw it as a critical inflection point during which she had to make clear the importance of evaluation and research to the foundation, and as an opening to define a more robust role for the evaluation function for the long term. She wanted to ramp up the use of outside commissioned evaluation as a way to inform the foundation's thinking, while ramping down the use of measurable results.

"Measurable results literally didn't tell us anything about our impact," Price said. "It told us nothing about was actually happening or why. It only told us how many people were getting served."

Given the fact that her boss had developed and was committed to the measurable results approach, Price needed to think through how to make it feel to executives that the new way of working required new ways of thinking about and doing evaluation. Could she keep the spirit and language of the measurable results approach while still making a substantial shift in what the organization believed evaluation could offer?

"We said very clearly that there is value in these measurable results. It's just not quite the value you thought it was," Price said. "We're going to keep that as a core component of the model. But we're going to be really clear on what it is …helping us understand. We would have liked to change measurable results more than we were able to in those early days. But we tried not to push too hard and too fast because folks felt comfortable with how they looked. And so we took a steadier, longer glide path to make changes to measurable results."

The evaluation team decided to keep measurable results while rapidly expanding investments in external evaluations. Leading up to her first budget proposal in 2014, Price shared her new vision and reasoning with the executive team to make sure they were on board with her ideas. In that proposal, which she called Eval 2.0, she asked for a budget increase from $185,000 in the previous year to $411,000 for the next year to fund six evaluation projects. It turned out to be the most detailed budget rationale she would ever need to present because the board and leadership were quickly convinced that the evaluations were worth the investment. The budget steadily increased in the following years, rising to $1.78 million in 2016 and funding 21 external evaluations.

These evaluations aimed to respond to the foundation's needs at the time as well as to demonstrate new ways that external evaluations could bring value. For example, the team commissioned short-cycle six-month retrospective studies on bodies of work that the foundation planned to continue and that staff could use in making strategic decisions (more on this later in the case). The evaluation team worked to show staff how evaluation could be useful by asking staff what their next strategic decision was and commissioning evaluations to help answer that question. Additionally, the evaluation team commissioned evaluations on complex approaches such as collective impact, and embedded evaluation from the start of new work.

Price and her team also wanted to make headway on helping staff ask more critical questions. At the time (and befitting how the foundation had historically envisioned that its role in change was selecting and supporting the individual organizations most likely to deliver on the measurable results), staff questions tended to focus on the capacity, stability, and leadership strength of individual organizations. In the shift to strategic philanthropy, Price wanted evaluative questions to focus instead on *portfolios* of work.

Although they had already begun to make headway on improving strategic thinking and supporting learning, this vision for the role of evaluation within the foundation would require more significant changes in how other foundation staff perceived the evaluation team's competencies, value, and authority.

Testing the boundaries of evaluation and research's authority

As strategic grantmaking portfolios were being formed in 2013 and 2014, foundation executives set formal criteria against which proposed portfolios would be judged. One of only a handful of criteria was the expectation that portfolios be evaluable, and the research and evaluation department was given the authority to make that judgment. This gave the department credibility as a partner at the table, Price said. When the evaluation team pushed back on something with program staff, they now had more authority to do so.

Sometimes the evaluation team determined that a portfolio was not clear enough about what it was trying to accomplish—or about the core mechanisms for change that they would be testing through their work—to be evaluable. But occasionally leadership approved a portfolio regardless of its evaluability. In these cases, the evaluation team made the choice to step back and not be too demanding about whether a portfolio was evaluable, so that they could preserve some of their relational capital for the longer-term change they were trying to make.

Program staff were under time pressure to get new funding opportunities launched and the evaluation team recognized that their insistence on making sure every piece of work was evaluable would have held up work and irritated the teams. Instead, the team typically declined to fund evaluations of the work, even when program staff requested them, because they felt no actionable information could come out of an evaluation. Price said that the team had to balance their newfound authority with the reality of the pressure program staff faced to get grants out the door in time to meet payout requirements.

In the summer of 2016, when the evaluation and research team were still feeling out their authority, they had a chance to test the boundaries of their influence with a new funding opportunity called the Health Insurance Literacy Initiative. This was one of the initiatives that came out of the strategic refresh and was an effort to become more strategic in the foundation's grantmaking.

Following passage of the Affordable Care Act (ACA) and the expansion of insurance coverage, staff were hearing from their coverage grantees that the next step was to make sure that people actually got access to health care services. People needed to better understand the benefits offered through the ACA and other plans so they could make use of health insurance policies, get health services like preventive care, and ultimately have better health outcomes.

Disagreements on what can be evaluated

Kyle Sargent, the program officer who was developing the initiative, wanted to use the funding initiative to learn about the comparative effectiveness of different approaches to increasing health literacy without putting too many parameters on the types of projects that would be funded. Chau was working closely with Sargent and began to have serious concerns about the initiative as it became more formalized.

Chau felt that the whole concept of health insurance literacy was more complex than it initially seemed, in part because of rapid changes to the insurance market and health care services at both the local and state level after the ACA. From the evaluation team's point of view, the initiative was conceptualized so broadly that it became nearly impossible to learn anything about effectiveness. In Chau's view, for the portfolio to be evaluable it needed to be much more specific in its articulation of what they were trying to accomplish and what range of actions would be tested. If anything goes, is there really a way to test and improve your thinking and work? The structure of the portfolio was such that the evaluation team believed that the work wouldn't actually enable the testing of any specific hypotheses about how to increase health literacy. To Chau, the way that the initiative was being proposed felt like a hold-over from the "responsive grantmaking" days, where most any definition of success or any approach could be included.

"Kyle had a lot of questions about what works to get people to understand and use their health insurance, but the way the initiative was written he was trying to respond to so many different needs and provide so little structure that there were no hypotheses attached to it," Chau said. "We weren't testing any models of what works. We were just going to fund a whole bunch of things."

At best, it would help the foundation learn "what's out there"—something that the evaluation team felt could be learned much more efficiently and cost-effectively through a straightforward landscape scan or convening.

For his part, Sargent felt that if the initiative was crafted too narrowly (in the form of clear boundaries around the kinds of approaches it would support and test), it could limit the foundation's learning about innovative approaches out in the community.

Wrestling with different assumptions about what it takes to learn

"I'd say it was the difference between wanting in my mind to meet the community where they're at in a messy world without getting too siloed and [the evaluation department] being more forceful about wanting it to be more targeted on how we measured and defined impact," Sargent said. "The intent of Health Insurance Literacy was to capture and learn as much as we could from it. My goal—and I think the goal the community had argued for—was to test some practices that were in the field or develop new practices as we had done with our Community Outreach and Enrollment Initiative… If we get too specific we're going to rule out some activities that might give us some data…. It’s that creative tension of how narrow do you have to be for internal purposes for measurement versus how broad do you need to be to meet the community’s needs.”

Sargent said he hoped the initiative could have communities identify what health insurance literacy is in Colorado and the tools and interventions that can lead to behavior change, meaning more enrollment and appropriate use of benefits. The idea was to identify through the application process areas of interest and then build and evaluate a portfolio that could inform the foundation on those questions.

Chau and Sargent, whose offices sat next to each other, went back and forth many times about why each thought their point of view about how to structure the portfolio to maximize learning made the most sense. In the meantime, the deadline for the draft initiative to be sent up for review and approval to the foundation’s executive leadership was rapidly approaching.

Chau shared her concerns with Price, who debated at length with Chau about what to do. Despite the requirement that portfolios be “evaluable,” it was not clear that Price had any authority to stop the initiative from going through as proposed. In his memo for this funding opportunity, Sargent had explicitly said that this was a funding opportunity to learn how health insurance literacy was linked to health and had proposed that $200,000 of the available funding be allocated to evaluation activities.

Worrying that this could be “an enormous waste of money”

But when Price and Chau looked at the memo, they completely disagreed that this approach would lead to any understanding of what works and for whom. Price said that at this point in her work at the foundation, she was getting excited that staff were starting to think about learning in a more systematic way. But she was also frustrated that they were setting up strategy without considering how the structure and approach could make evaluation and learning possible.

“There was the promise that we were going to learn how health insurance literacy is linked to health outcomes,” Price said. “You absolutely cannot learn that by giving grantees an enormous amount of latitude and not having any structured evaluation that is measuring health outcomes and measuring what kind of health insurance literacy is happening.”

She said sending up this funding opportunity for approval “as is” felt like she was telling the executives that she was on board with what the funding memo promised about learning when she in fact was not. Price said the $3 million price tag for the initiative especially raised her concerns. Price felt she could only give her stamp of approval if she could describe how this strategy had a chance of creating change, “rather than just being an enormous waste of money.”

Chau, meanwhile, who had worked with Price for about two years at this point remembers, “I think it was the first time I saw Kelci genuinely confused and upset because she’s a pretty cool natured person. It was the first time I saw her body language that she was like ‘what the hell is going on?’”

While her authority and role in this decision making was still not clear, Price had a philosophy that she was following.

"My general philosophy was to simply start acting as though we had permission to make decisions about things—rather than trying to do the rounds and get buy-in and ask permission," she said. "When there is a vacuum, you can step into the vacuum with something and people will generally adopt it. It's more challenging if somebody already owns that turf and you're trying to get them to change the turf somehow. But we had a lot of vacuums here… just like fallow fields. And so as long as the thing I was introducing had obvious value to people, they were willing to start doing it. I do think that there is a tension in evaluation writ large—and I experienced this with our evaluation consultants and I've experienced it my whole life as an evaluator—this whole idea of listening to the client and then responding to *whatever* the client asks. It puts evaluators in this very reactive stance and I see the same things happening within foundations."

Deciding to push the envelope

Price decided to speak with Sargent's boss, Erica Snow, about her concerns. Snow had recently been promoted as an interim portfolio director as part of the re-shuffle after the CEO and several vice presidents left. Snow used to be Sargent's peer and was still getting her sea legs on both being his supervisor as well as taking on her new responsibilities. While Price felt strongly about her concerns, she also felt uncomfortable about initiating this conversation. She noted it was not common for peers to go to one another and hash out disagreements (Chau and Sargent's experience notwithstanding), let alone go to someone who was in a different position in the hierarchy.

"Our culture did not support candid conversations," Price said. "The preservation of relationships was core. People felt like candid conversations threatened relationships. So, we had a nicey-nice culture where you didn't say candid things to the other person because it was interpreted on a personal level, like [questioning a strategy] was an attack on *you*…. [Additionally] there was this outstanding question that we'd never as an organization tackled, which is *what authority does the evaluation team have to put up a stop sign*?"

After hearing Price out, Snow agreed to have the evaluation and learning team host a "design blitz" session to take another look at the funding opportunity. As a result of this session, Sargent removed one of the three strands in his proposal that was focused on systems change. The other strands were focused on messaging and connecting networks already being accessed by individuals and families, both of which had slightly clearer boundaries and thus potentially testable (if only implied) hypotheses. Price said that was still not the level of change she hoped Sargent would make to the strategy, but the design blitz seemed to give others in the room permission to voice their own uncertainties about the proposed initiative.

"Some people said, 'I'm really confused about this too' and that had never been said in previous meetings," Price said. "It was the whole candid conversation issue. It was gratifying to discover that because I felt like I was out on a limb, kind of alone doing this. It felt kind of lonely to do this in an organization and felt very uncomfortable in the new roles that we were carving out. It was very gratifying to discover that we had, in fact, created a forum where other folks could acknowledge that they were also confused and concerned about this strategy and didn't understand it."

For his part, Sargent said he was surprised that Price went to Snow with her concerns before speaking with him directly. He and Chau had recently collaborated successfully on an evaluation of the foundation's initiative on outreach and enrollment that he felt had gone well.

“We were trying to change the culture, but we still had this very collegial, cross-functional approach,” Sargent said. “It might not have been the most efficient way of doing things, but it had proven successful. Kelci going to Erica and making that decision was different than the way that we thought we were going to be implementing it, at least from my perspective where program staff truly held that final decision. And I thought that’s not the way I would do things as a partner. So I think the impact was one of frustration…. It was not a ‘hey, let’s sit down and talk about this with Kyle, who’s leading it.’ It was Kelci going to Erica, but she didn’t come either to me or through Yen to say ‘hey, you know, we’ve got some problems and some issues with this.’”

Sargent continued, “When the decision was made that we were going to be changing our approach, eval basically took over and their concept was that they wanted to do the design blitz. So the way it happened, I would argue it should not have been that way in a true collaboration. But given the fact that it demonstrated how much control eval would have over philanthropy in the decision-making process, I’m not sure the result would have been any different in the end. I report to Erica, so if Kelci and Erica decide that we’re going to do something different, then we do something different. You know, that’s life.”

New CEO and Another New Direction

In September 2015, just as the staff were beginning to become accustomed to their new approach to grantmaking, Karen McNeil-Miller took the helm as the new CEO of the Colorado Health Foundation. The former CEO of the Kate B. Reynolds Charitable Trust in North Carolina, McNeil-Miller brought a fundamentally new vision to the foundation around organizational values, community engagement, roles of program officers, and diversity, equity, and inclusion.

“I wanted to bring to the foundation a greater focus on the external environment rather than being so internally focused and driven, which the foundation had the reputation for nationally at the time,” McNeil-Miller said. “It had the reputation of being an ivory tower, coming out with proclamations of what was best and acting as the experts. Our job isn’t to think great ideas and just put them out and decide if they work or not. If our mission is to improve the health of Coloradans, we need to understand what that means to Coloradans, for Coloradans. We knew a lot about organizations, we knew a lot about approaches, but our view of community and the people of Colorado were that we would learn about them and then tell them what they needed to do as opposed to we will learn from them and with them about what their needs are.”

A new ally who finds the evaluation department frustrated

McNeil-Miller is a big proponent of learning and evaluation, which provided the team their biggest ally to date. But when the new CEO took her measure of the culture, she found that the evaluation department had been facing some uphill struggles. She described the foundation as a conflict-avoiding organization, a place where it was not fun to work, and where people played things safe because if they made a mistake by taking a risk on a grant or proposing new policy work that did not pan out, they worried they would get penalized for it.

At the same time, she said that the philanthropy department was “the sun” around which all of the other departments, including evaluation, orbited. McNeil-Miller felt that the foundation had a larger mission that should include evaluation, communications, and policy as equal partners in working toward creating an impact. All of these issues affected Price's department, McNeil-Miller said.

"The evaluation department existed to serve philanthropy's needs and I didn't like that perspective," McNeil-Miller said. "I found Kelci and her team frustrated. I saw Kelci struggling with the fact that they weren't listened to. I recognized the stress and distress that she and her team were feeling of not feeling valued or listened to and their expertise was kind of dismissed. They were just kind of 'fetch-its' as opposed to equal colleagues participating in the work.

We would have Strategy A and say 'we want you to measure this' and Kelci and her team would say, 'okay, we can measure this but that is not going to tell you whether that strategy is working or not.' But because 'the sun' ruled, they just ended up measuring things even though they knew that they were not really the things to assess for impact or effectiveness."

Negotiating the role of evaluation staff in the midst of change

Taryn Fort, the foundation's communications director, has also observed the evaluation department wrestle with its role.

"The learning and evaluation team clearly bring the talent and abilities needed to drive strategy on their own," Fort said. "They often present different ways of thinking about what strategy is or isn't. They're also strong facilitators because they are unafraid to ask questions that probe us to think further or explore a concept more deeply. Yet, I see them staying in a holding pattern at times, while waiting for others in the organization to make strategic decisions so they can support them. And other times, they end up driving strategy because others don't step up or know how to do it. Learning and evaluation are gap fillers, so if a solution isn't found or explored, they often end up leading that discussion or effort. I see that creating time management issues for them, at times. This can also create tension and expectations that someone from learning/evaluation must lead strategy development, putting some pressure on them to be the deliverer or driver of strategy, especially when we need others in the organization to also bring strategic thinking and practice to the table."

Snow, a portfolio director who oversees senior program officers, said that from her side the questions that the evaluation department raises, while often good, may feel like another roadblock when staff are under time pressure to get a big backlog of grants through, especially as deadlines are fast approaching.

"There's often a good six weeks where there is a huge amount of grants moving through the system, and I think the tension is that program staff just want to move the grant forward and can experience it as burdensome to have the pushback," Snow said. "That can happen particularly at the end of the year when we're the only ones held to meeting our 5 percent payout."

For some, evaluation department is perceived as critics rather than partners

Price and Chau's approach may also be a barrier for some, said Snow. Program staff have shared that at times they do not experience the learning and evaluation team as partners in strategy development, but instead as critics. At times, they feel their learning and evaluation peers do not seek to understand or empathize with what program staff are experiencing in the field, which may lead to program officers being too quick to dismiss them. She said that it might be helpful if Price and Chau positioned themselves more as actively engaged partners at the table who have their own struggles and concerns, rather than taking a more detached, observational stance.

“Senior program officers say they don’t always feel that direct partnership and connection with Kelci and Yen,” Snow said. “Some staff have expressed feeling unsure when to go and talk to them or not. They are uncertain about what is in and out of bounds in terms of evaluation staff roles and responsibilities. Folks say ‘I go there and feel foolish, like I’m not doing my job right and then I don’t want to ask [them any questions]’.”

Karen McNeil-Miller, meanwhile, said “I think Kelci would tell you that one of the things she is working on is her influencing strategies. When Kelci gets in the grip and things aren’t going well, she digs into the facts and the research and continually telling people why this won’t work. She’s understanding that that’s not a very good influencing strategy.”

Instead, Price should try and understand how others came to their decision including their thought process, McNeil-Miller suggested. “She could have asked more questions rather than making more declarative statements and engage them in conversation, as opposed to them feeling like she is trying to tell them how to do their job.”

But others respond well to Price and Chau’s challenging questions. In 2016, Jehan Benton-Clark came to the foundation as a portfolio director and immediately saw the value of the learning and evaluation department to her thinking. She said she enjoys when people ask her hard questions that challenge and push her. When Benton-Clark started, she felt as if the evaluation department was poised to take a more active role, but needed the “cover” of McNeil-Miller’s support and approach, which is to ask a lot of questions and see all of the foundation’s work as an opportunity to learn and improve.

Contending with a long-standing culture of perfection

Benton-Clark, meanwhile, said that another factor that the evaluation team must contend with is a long-standing culture of perfection at the foundation. That culture, which she says is improving, has put a lot of pressure on staff. Some can feel that if they are being asked hard questions by the evaluation team, it’s not from the perspective of improving decision making but rather an attack on them or the work they are doing.

In 2017, Nick Stuber joined the learning and evaluation department as a learning and evaluation officer. Stuber had extensive experience carrying out evaluations, had done some strategic planning work and facilitation work, and said he was eager to synthesize that knowledge and deepen it at the foundation. As Stuber got to know a bit about the foundation and the organizational dynamics, he said he had an approach in mind in working with program officers.

“I cast aside [my title] right away and sat down with program officers as a thought partner, as an internal consultant,” Stuber said. “I said that I’m really here just to be someone you can bounce ideas off, and I can help you think through what those are and get clear on decisions you need to make around your body of work. I found that the most helpful role was to really not use any of the formal language or terminology, but to frame it as I can be another colleague who’s here, who can help you because the senior program officers are the ones really leading the body of work. Once I understood their role, I could say I'm a colleague who can help you think through all that so you don't have to feel like it's all on you to produce this work."

Benton-Clark said that Stuber soon established a connection with program officers, recalling, "He was able to build a good rapport with his colleagues and other departments, and he's also willing to push and challenge. I think [program officers] started to see him as a thought partner pretty quickly."

Price and Chau said that they have established strong relationships with staff at all levels of the foundation in their work across time, though it is not easy to manage the tension inherent in evaluation's role as both critical friends and supportive colleagues. In doing so, they each have taken a particular approach to their role. From Chau's point of view, the evaluation department's role is to ask tough questions.

"This [work on Health Insurance Literacy] did raise the question with Kelci about what is our role," Chau said. "And I remember she said to me 'philanthropy [staff] will always have the final say, but our job is to raise questions.' I've told Kelci many times it's like if you see a friend about to slam into the wall wouldn't you tell them not to? And she says, well, you know if they are determined to ram themselves into the wall maybe what we do is provide a mattress."

"I've always said that I think there needs to be healthy tension between evaluation and philanthropy because we're the mirror," Chau continued. "Sometimes the mirror shows you not in a very great light and we need to do be able to do that. If we are a true mirror of who you are then it shouldn't always be pretty. If we always agree, then there's something wrong…. For a while, we were just not pushing back on philanthropy hard enough…. If we are not asking critical questions and having hard conversations, then we are not doing our job properly."

McNeil-Miller, meanwhile, has set a path for the foundation where she establishes broad goals and then all the departments align with those goals, rather than having the philanthropy department as the "sun" around which everything else orbits. This approach gives a new importance and independence to the learning and evaluation department, as it does for the policy and communications departments.

New direction for the foundation and roles for program officers

In 2017, the Colorado Health Foundation announced a new strategic direction for its work—the third in five years. Its new focus was informed by data that had been collected by the evaluation team from communities across the state, a new philosophy about community engagement, and a new focus on putting health equity at the center of everything the foundation does. The foundation announced it would adopt a wider view of health and start working more directly in local communities. The foundation expanded to four focus areas: maintain healthy bodies, nurture healthy minds, strengthen community health, and champion health equity. At the core of this work was a more deliberate focus on improving the health of low-income Coloradans in underserved communities, particularly rural areas, being informed by community, and doing everything with the intent of creating health equity.

As part of this transition, the roles of program officers shifted dramatically yet again. Rather than topical experts engaging in strategy design and grantee selection from their offices in Denver, they are expected to engage with communities with a "cultivation" mindset that promises a different approach to developing relationships and potential grantees. Staff are expected to be out in the field a minimum of 40 percent of their time, engaging with a variety of people from CEOs of organizations, to community advocates, to people working in schools, health settings, and others, some of whom would *not* be in the pool of potential grantees. They are expected to gather information and perspectives on community and system needs, opportunities, and dynamics, and then synthesize that information. This deeper level of community engagement, in theory, will increase the foundation's understanding of the larger context, and position them to better respond to factors impacting health and equity at the community level.

Program officers have been re-assigned to cover new priorities and geographic regions, with almost all of them leaving behind long-standing relationships, grantees, and work in a particular issue area. Instead, they are expected to seek out new views and build relationships with a set of organizations that they may have never interacted with before, on topics they may know nothing about. The foundation supported this change with a broad set of practices including professional development, monthly reflective time together to debrief the new ways of working, and opportunities to learn about new issue areas.

They are also expected to do more than rely on grant applications as the mechanism for identifying potential grantees, instead exploring through their work in communities whether and how a potential grantee serves a community in a way that aligns with the foundation's interest and values and the needs of that community. In addition, program officers have been asked to approach their work within communities, and their selection of grantees, with a commitment to health equity—a value that the foundation is working to understand and articulate for the first time.

By taking on this role, the foundation is hypothesizing that it can have a deeper impact because its dollars will be more directly targeted to individuals experiencing the most inequity in health and health care, and will be used in more contextually-specific ways that capitalize on existing energy and community-defined needs and desires.[5]

Enormous change in expectations for staff

This and other new skills required for program officers represented an enormous change in expectations and assumptions about what it meant to perform well and what program staff were now accountable for. In the responsive grantmaking era, program officers' performance had been judged largely on their ability to identify and select grantees from a pool of proposals that could roll up to the 12 measurable results. In its brief stint with the form of strategic grantmaking that resulted from the first strategic refresh, program officers were judged largely on their ability to develop and articulate funding opportunities and then select grantees who could deliver the goals and approaches the foundation outlined.

Expectations of program officers became fairly expansive, emphasizing trust-building, brokering connections, supporting strategic analysis and problem solving within communities, and seeding projects and activities that hopefully lead to bigger, more strategic health improvement projects driven by local groups. Now, program officers were expected to read local contexts and dynamics

[5] See the foundation's [public articulation of its program staff role](https://coloradohealth.org/sites/default/files/documents/2018-06/IMPACT_PracticeModel_March2018.pdf) at https://coloradohealth.org/sites/default/files/documents/2018-06/IMPACT_PracticeModel_March2018.pdf

effectively, and see strategic opportunities to support locally-driven action that was in alignment with the foundation's equity commitment and focus area goals. In addition to gathering information through their community engagement, they now became the "quarterback" of a strategic approach. Rather than operating as the individual "owner" of a strategy that other departments support, program officers led an entire cross-functional team, including philanthropy (programs), communications, evaluation, and policy, that took on that responsibility.

This last change relieved some of the pressure program officers felt in bearing the full responsibility for strategy design, as well as some of the tensions from the evaluation department asking hard questions. But, overall, many program officers did not feel prepared for or interested in their new role and a number left the foundation.

Learning comes to the fore again

With this shift in role, and routine messaging from new leadership that uncertainty, "failure," and learning are an expected part of the work, the foundation now talks about itself as testing a more adaptive, iterative strategic process. As a result, program officers began asking the evaluation department—now renamed as the learning and evaluation department—to help them with rapid cycle learning. The evaluation team has taken those requests and expanded them to set up and oversee comprehensive learning plans over time.

"We're starting to get these requests from many of the teams because they're recognizing that they need to have different conversations when they're vetting grants," Price said. "And so they want us to set up conversations for grant review that are fundamentally different than what they have done before, which requires analysis of the grants that have come in and the structuring of conversations. All of a sudden there's not only the strategy-level learning work, but there's the question of 'okay then how does that strategic conversation trickle into all of those things we're doing around a strategy like selecting grantees?' We've never really been invited into that door. It's a door we've been looking at for years—how do we make those [grant vetting conversations] more strategic."

In addition to the new roles required of program officers, the foundation's new commitment to equity also meant that the learning and evaluation department needed to re-focus its work toward equitable evaluation (discussed later in the case).

Commissioning an evaluation in the midst of great change

Amidst all the uncertainty, the evaluation department continued to commission external evaluations. Two months after McNeil-Miller started at the foundation, in November 2015, the evaluation department launched an evaluation of a funding opportunity that had come out of the strategic refresh under the previous CEO.

During the refresh, the evaluation department aimed to link its work to strategy and demonstrate the usefulness of external evaluations by more explicitly tying evaluations to program officers' next strategic decision. One of the funding opportunities that came out of the foundation's first strategic refresh was an initiative called Healthy Eating Active Living (HEAL) Advocacy, aimed at building the capacity of organizations involved in advocacy related to healthy eating and active living. The program had three streams: capacity building, leadership support, and policy initiatives.

"We had made the decision to take this on as a strategy… because we felt the lack of strong advocates for the healthy eating and active living issue in Colorado, and that had been a limiting factor in producing policy changes that we thought would help with the overall strategic goal of the foundation at the time, which was reducing childhood obesity," said Kyle Legleiter, senior director of policy advocacy at the foundation.

Testing the idea of linking evaluation to specific decisions

The field of advocacy organizations working on this issue was diffuse, in part because many of the policy changes that needed to be made would happen through policy decisions at the local level rather than in the state legislature or other state-level policymaking bodies. The foundation wanted to learn how to get better at supporting grantees to do the very specific kind of strategies required to influence policy at this level.

"How do we start from a place of not a lot of capacity, and grow that capacity not just for capacity's sake, but then also for the sake of ultimately influencing policy outcomes and people's opportunities to engage in healthy eating and active living in our state?" Legleiter wanted to know.

The evaluation department commissioned the evaluation firm ORS Impact to carry out an evaluation of the HEAL initiative. ORS developed evaluation questions in response to the foundation's expressed aims for the HEAL evaluation, shared in the request for proposals and in subsequent design meetings. The evaluation questions included:

1) What outcomes are associated with advocacy leadership grantmaking? How strong are they?
2) What outcomes are associated with organizational capacity building? How strong are they?
3) To what extent and how are connections being built among advocates?
4) How are grantee and community members' skills and connections being deployed to advocate for HEAL? For social determinants of health?
5) To what extent are advocates forming an effective field and how is it evolving?
6) What factors facilitate or create barriers to advancing key outcomes?
7) What outcomes appear to have highest value in building a base of support, strengthening HEAL advocacy, and advancing HEAL-related policies?
8) What impact have Colorado Health Foundation grantees and other HEAL advocates had on policy change and system reforms in Colorado?

When ORS Impact began this work in November 2015, the foundation was in the midst of a number of changes. While the strategy refresh had set a new direction for the organization, foundation staff were still figuring out exactly what that meant. In the meantime, McNeil-Miller had just arrived as the new CEO and many staff assumed that she would chart an entirely different path for the foundation. This meant that the evaluation team was hard-pressed to identify what strategic decisions would likely still be on the table by the time the evaluation produced any findings.

Understanding this, Price and ORS aimed the evaluation at learning what they could from the current strategy about how to build a field of advocates who could affect local policy, often in communities where very little capacity existed, but also left room in the scope of the work each year to meet emergent needs.

Early questions about the initiative

About a year into the work, after doing some initial data collection to understand the universe of grantees, their focus, and the strategies they were using, ORS Impact staff raised a concern with the evaluation team. The program officer in charge of the initiative was making grants that did not seem to be in sync with the strategic intent. She was funding a wide variety of initiatives that broadly addressed social determinants of health, such as economic security, housing, transportation, and immigration issues, even though the funding opportunity as approved by the board was written much more narrowly to focus on local-level advocacy for specific issues of healthy eating and active living.

The funding opportunity had been designed to allow for funding social determinants of health, knowing that there were few advocates focused specifically on HEAL, but a strong argument was supposed to be made that there was a direct connection between the social determinant and HEAL outcomes. The ORS team judged that the grants being made did not necessarily have this strong connection.

The questions the evaluation was originally designed to answer made little sense given the gap between the intended strategy and what was happening on the ground. ORS evaluators Anne Gienapp and Carlyn Orians could not find much evidence that grantees even viewed themselves as working on healthy eating and active living to begin with, much less that they were working together to build a field. Given this, the evaluators were, understandably, unable to detect much evidence of the impact that had been articulated in the original strategy. ORS staff felt this was a big enough concern to bring to Price's attention. They did this after a lot of internal debate. This was their first engagement with the foundation, and they would be calling out the work of the program officer they were working with. By talking to Price, they would be going over the head of both that program officer and the evaluation officer with whom they directly worked and from whom they had struggled to get insights on changes taking place at the foundation. It felt risky. But they also felt the issues were so critical that they had to raise them with Price.

They arranged a call with Price in January 2017. Price remembers that Gienapp and Orians asked what role evaluation could best play moving forward, given that the work on the ground didn't match the intended strategy. About a month later, ORS Impact and a number of foundation staff met to discuss the evaluator's concerns and consider next steps. Among those participating were people in charge of implementing the foundation's new strategy, which ORS Impact hoped would help give them a better understanding of the changes taking place at the foundation.

"I was thrilled that they came forward," Price said. "What we always tell our staff is that part of our role and that of external evaluators is to surface disconnects in what we are doing. It is a great example of them raising the issue of lack of alignment and giving us as an organization a chance to decide if we needed to change our implementation or our goals."

As it turned out, ORS Impact's concerns did not lead to any shifts in the HEAL initiative. At the time, the foundation was starting to implement the entirely new strategic vision from McNeil-Miller, and, as part of that, moving toward more work in social determinants of health and work in rural communities. Making any big changes did not seem to be worthwhile as it already seemed clear that the initiative would sunset. In addition, the program officer left the foundation when the new roles for the program staff were announced, and there were several additional transitions in program staff assigned to oversee this funding opportunity over time.

A new direction and more questions

However, the meeting did yield a new direction for the evaluation. After making clear that the original plan for the evaluation was unlikely to yield helpful findings, ORS and foundation staff discussed other questions that the foundation was wrestling with where the evaluation team might be able to provide some insights. Given what they had heard about the foundation's emerging focus on local communities, including rural areas, ORS suggested they research advocacy capacity building in rural communities based on the handful of rural communities represented in the HEAL initiative. This would hopefully help the foundation learn something useful for its future work in rural areas.

The instructions from the foundation to ORS were to use their remaining Phase 2 contract dollars to switch to a case study approach focused on the rural HEAL communities. The team was concerned about having a sufficient number of cases to be confident that their findings were generalizable. Only three of the 12 HEAL grantees were in rural communities, with one additional community that contained both rural and suburban zones. In an effort to increase their confidence in the conclusions, ORS focused on these four communities.

Initially, ORS staff felt re-energized and that they were on the right track after a year or so of feeling that their work was not going to yield useful insights. But when they shared initial findings with the foundation staff, including policy officers, they did not get much of a response.

"[The reaction] was a little lackluster," Gienapp said. "It didn't feel like totally a lead balloon but it really did not feel like it excited anyone. It was a bit of a bummer. It was a letdown for sure."

Price felt this way, too.

One of the pieces of feedback from foundation staff was that the way ORS was casting its findings did not include an equity lens in the way the foundation was coming to understand equity, or consider what it means to focus on structural conditions that drive inequity. ORS evaluators said that, while they knew that the new CEO was deeply committed to an equity approach, they did not know where this new focus was taking them or how the foundation was thinking about structural conditions that drive inequity.

When Legleiter reviewed the report, he remembers being concerned that one of the communities featured came from a context that Coloradans would perceive as urban or suburban. This particular community is located in the Denver metropolitan area. The credibility of the conclusions overall would be questioned once people saw this.

"To me, that was one of the misfires of that report," he said. "If we wanted to understand how advocacy capacity might be different in a rural context and if part of the conclusions were from a context that was not actually rural, that prompted some suspicion about the…findings that they had generalized across cases. I expected a lot of criticism if we put it out there… I suspect that [ORS] may have made that decision to include a community that's in fact not rural under some duress, feeling some pressure [to have] enough case examples that it wasn't overgeneralizing from just two case studies or something like that."

Legleiter also noted that he began his career in rural Colorado in one of the other communities that was profiled in the study. For him, the findings in the report were not surprising or groundbreaking. What would have been more helpful to him as a funder beginning to think about drivers of inequity would be to dig into some of the structural challenges in rural communities and why there is more limited advocacy capacity in those communities.

He recalls that the report synthesized reflections from the organizations already doing advocacy work with grants from the foundation (those from whom the evaluator had already been collecting data). But it did not focus on the strategic questions that were the most pressing for him to understand, such as what advocacy can look like for a community without many 501(c)(3) organizations at all, or without ones that have experience doing more than direct service work, or that have permanent staff. How could the foundation support advocacy in these contexts where basic nonprofit infrastructure did not exist and/or was chronically underfunded—the conditions the foundation was most likely to encounter in its new rural engagement strategy?

Legleiter recalls, "It didn't really do anything to advance my thinking about those structural, underlying, contextual conditions."

Legleiter cannot remember if he was asked to weigh in on what kinds of questions might be useful to him.

Trying to understand the rapid changes at the foundation

Gienapp and Orians said that while they had a strong, collegial, and trusting relationship with the evaluation team, an ongoing challenge they struggled with was trying to discern and understand the rapid pace of change at the Colorado Health Foundation and how to apply it to their work. One reason they suggested the larger meeting in February 2017 was to get a clearer understanding of the changes that could inform their work, Orians said.

"In our day-to-day contacts at the foundation, the people we were getting information from weren't at the executive level, they were more at the program officer level," Orians said. "Every once in a while, we would get some piece of information that alluded to this larger change. But we were concerned that we weren't understanding well enough the nature of that challenge. [As a result] we didn't always have a clear direction for what would be most useful."

Added Gienapp, "In working with a lot of foundations, we've found there's a spectrum. There are foundations that share more information about changes that are happening that would be useful to an evaluator who is acting as a partner, and foundations where intentionally or not that doesn't happen and they keep things closer to the vest. At [the Colorado Health Foundation] I found with the evaluation officer, the program officer, and even to a certain extent Kelci, that they were on the end of the spectrum of keeping things closer to the vest. Repeated attempts by us did not necessarily yield a greater understanding of what was happening in the organization, which made it difficult to navigate as an external evaluator."

Both the foundation and the evaluator could have been more proactive

From Price's perspective, the foundation could have shared more with ORS *and* ORS could have pushed harder to understand the changes taking place and implications for their work.

"I took for granted that Anne and Carlyn understood the strategic changes and thinking that was happening at the foundation—believing they were either getting it from my team members or the program officers," Price said. "Through a combination of staff transitions overseeing the initiatives—three different program officers—and the fact that apparently my team member did not discuss [the changes] with them in appropriate depth, I only later learned that they did not feel like they understood the shifts. Part of that was definitely us."

"But I think ORS could have been more mindful of deliberately connecting with us to understand what was happening, and to be very actively updating their approach to evaluation to address this," Price continued. "They knew that organizational shifts were underway, but I don't think that they really intentionally reached out to have a discussion. Each time there was a transition of a program officer would have been the perfect opportunity to regroup. And if they felt like they weren't getting that information from my staff member, they had a direct connection to me as well that they could have leveraged to get this information and have a discussion."

She noted that when something came up in the evaluation that piqued their interest, ORS routinely considered how that might be leveraged into future decision-making for the foundation. The challenge was that their timing was too late. Because internal change conversations were happening outside the purview of the evaluation consultants—and sometimes even the program teams—the ORS team did not have visibility into the fact that the decisions had already been made and new work was already happening. They could not propose an adaptation to the evaluation that would add real value because they could not see what types of strategic questions the foundation had or how their strand of work connected to other work outside their scope.

While the rural case study was not as impactful as ORS had hoped it would be, the evaluation team undertook the final phase of the evaluation, at Price's request, which continued evaluation activities specifically related to the leadership development and capacity building streams of the funding initiative, and which was submitted in February 2019. Orians noted that one piece of feedback the team received was that many of the recommendations they made were already part of the foundation's latest strategy, which was again new information to ORS.

Ongoing challenge with evaluations

Beyond the specific challenges that ORS faced with keeping pace with the foundation's many transitions, Price said that she has found an ongoing challenge with other evaluators in getting helpful findings. She estimates that only about 55 percent of the evaluations she commissioned are ultimately helpful to the foundation's strategic decision making.

One of the key dilemmas is that from her perspective, evaluators tend to provide a lot of data, almost like a data dump, without providing meaning or real synthesis that links data and insights to the strategic choices on the table.

"We've had evaluation teams come back with data and analyses and we're like, we don't know what we're supposed to get from this. And they're like, 'but its data!' as though any data is going to be useful and valuable," Price said. "And so I keep discovering that some evaluation teams just think that if they bring anything, we will somehow figure out how to make strategic sense of this and then use it in our decisions."

And at the core, many evaluators do not distinguish between a good strategic question versus a good evaluation question, Price said, and then figure out how the two connect. She wants evaluations that clearly test and then help revise and refine the foundation's thinking about how they can make change happen.

Part of the issue is that the *level* evaluators are looking at does not match the level of the foundation's work. At the program or strategy level, evaluators look at questions such as how well is the implementation going and is the foundation getting the results they hoped for from the implementation. Price refers to these as "single loop" questions that are looking at the success or not of an initiative (Are we doing things right?) At the portfolio level where the foundation sees its "value add" beyond the work of individual grantees, evaluators need to also ask regardless of how well the implementation is going, is this the type of work that will lead to the outcome that the foundation wants or should they focus on another strategy entirely (Are we doing the right things?) These "double loop" questions become much more important for evaluators to answer when working at the portfolio level.

"The conversation we've been looking to have is essentially the loop from: this was your thinking about why do these things, this was what you intended to happen, what actually happened? And then what does that mean for how you need to revise *both* your thinking and your action?...Is it a theory or an implementation failure?" Price said.

The evaluation team experiencing its own challenge with learning loops

The evaluation department had fallen into its own trap, too, of focusing on the wrong "level" of question in its own learning practice. In the spirit of "what's good for the goose is good for the gander," the evaluation department held itself accountable for completing and practicing the same routines it asked of program staff and had been routinely using emergent learning tools, such as After Action Reviews, to improve their own learning. At the outset of their experimentation with emergent learning, they had focused their questions most relevant to the immediate on-the-fly challenges they were facing as a team, e.g., "What do we now know about how effective it is to embed a theory of change into the planning documents for initiatives?"

At first, the approach proved as energizing and effective for the evaluation team as it had for the program staff. They gained insights about how to improve some processes where the first iteration was not producing the impact they had wanted, and abandoned some others they discovered were much less valuable than they had anticipated (such as requiring a formal theory of change process for each program team). However, after the first 18 months of using the approach, as the foundation's new strategic approach began to "settle," the evaluation team began to feel that its value for their own learning was slipping. Their routine reflections on how things were progressing stopped feeling like they were adding real value to their strategic decision-making about how to improve the foundation's use of evidence to inform the work. Price wrote:

> [Our] After Action Reviews had been strongly focused on ensuring that the new processes and tools we were implementing were "working right," but once these core practices had been put in place… we were no longer feeling the same urgency to focus on rapid-cycle improvement as we had during the launch phase in 2014. By 2015…the attention and interest of the team began to turn to larger questions, such as how program teams were integrating learning into their decision processes. This latter question reflects the bigger intended outcomes in our department's theory of change, and encompasses a whole host of activities and structures. Our AAR practice that focused on a single activity didn't allow us to address these types of questions, and we had not transitioned to using emergent learning to focus on our broader intended outcomes. I realized that we'd kept our emergent learning (EL) practice focused on activities for too long, and had not evolved our EL practice to include conversations about the broader outcomes we were seeking. This was leading to unstructured discussions about our bigger outcomes and was at the core of the issue that these conversations felt like 'admiring the problem.'

In short, the team recognized that they were no longer asking the questions focused on the right "level." In retrospect, Price noted,

> When our focus was on just getting things up and running in a fast-paced and uncertain environment, when we were still learning our way into our overall approach to evaluation for the organization, we benefitted from focusing on the question of 'are we doing things right'. When we achieved some stability of our core infrastructure for evaluation and could move beyond having to constantly be inventing things on the fly, the team's thinking could turn more broadly to what progress we were making towards broader outcomes [such as increased use of evidence to inform foundation decision making]. Our feeling of being stuck and not progressing in our own learning told us that we needed to change something substantive about our practice of EL, though we struggled for nearly a year to identify what was not working about our conversations.

The team began to articulate and explore what "double loop" questions *they* should be asking about their own work? Were their initial guesses about what it would take to increase the foundation's use of evidence to inform strategy bearing out? For example, their learning had been focused on things like how to successfully implement a theory of change approach. But perhaps now the more important thing to be asking, gathering evidence about, and then reflecting on was whether using a theory of change approach actually resulted in strategies that were more evidence based.

Having this experience themselves triggered a new focus among the evaluation team on whether they were helping program teams ask the right level of questions about their strategies. And how could they better set up evaluations to really test and refine teams' thinking and action at a level that would make a difference in the work?

New locally-focused work and questions about evaluation

In January 2018, as part of its broader focus on health equity, the foundation announced the launch of a new locally-based initiative to work in four rural communities around Colorado. The aim of this work is to help the foundation better understand where communities lack resources and connections, and to fill those gaps with homegrown solutions, according to Jehan Benton-Clark, who is leading the initiative. The ultimate goal is to help communities improve population health and reduce health inequities. The foundation aims to work in partnership with communities so that they experience strong, responsive, and inclusive institutions that enact policies and systems that promote health. Community members can use their power to engage, lead, and take action, and work together to address health-related challenges.

Benton-Clark started in 2016 as the foundation's new portfolio director overseeing the foundation's grantmaking efforts related to advocacy, community solutions, capacity building, and locally-focused work. From the start of her tenure at the foundation, she began to build a relationship with Price and Chau and make use of their services.

Early in the initiative, Chau began helping Benton-Clark by providing background on the foundation's previous place-based work and some thinking around data collection, which soon morphed into helping to develop a theory of change. During this period, Benton-Clark, Price, and Chau hired an external evaluation firm to do an evaluation and act as a thought partner. The firm developed an evaluation framework after extensive time spent with the program team to understand their thinking and work. But when Chau saw it, she felt it was missing a number of important pieces and, after a prolonged period of back-and-forth with the partner, decided to take over its development.

"Kelci and I have always talked about we're sometimes the worst and sometimes the best clients because we can actually *do* evaluation. So we push our contactors pretty hard on their methods and even the surveys and how they set up questions," Chau said.

As she dove into creating the framework, Chau had in her mind that, from her perspective, there have been few strong place-based evaluations. Chau wanted to create a framework for a learning evaluation that did what she had not seen in many other place-based evaluations: follow the storyline, assess the core hypotheses, and bring the learning back into the system to inform real-time decisions.

Differences in expectations and needs

As Chau was working on an increasingly elaborate plan, Benton-Clark was becoming increasingly impatient. From Benton-Clark's perception, the evaluation seemed to be focused so far only on learning without a connection to collecting external data that could describe what was happening on the ground. From Chau's perspective, it made sense the evaluators were not yet in the field because there was not much, if any, data to collect given that program officers were in the early stages of getting to know and build trust in the communities. But things came to head one day in August 2018, when Benton-Clark called Chau and told her she needed to start showing some type of data to Latham, McNeil-Miller, and the board, Chau remembers.

"I felt like we're still in the really new phase of this work and we don't expect change to happen in communities right away," Chau said. "So I thought I had a little more time and then Jehan was like '*no,* we need the evaluator to be out in communities collecting data *now*.' I go, 'on what, Jehan, on what? What is it that you want me to collect data on? Nothing is changing. Why is it this year, some arbitrary date, versus when we actually expect change to happen?'"

Benton-Clark replied that they had been doing some of this developmental work for two years so there should be some information to collect, Chau recalled. After that Chau worked with the evaluator to start collecting preliminary data about how the program officers were operationalizing their new roles and what early signs might be emerging regarding trust building and activation in the rural communities where they were working.

"My error was that I was trying to put together the long arc of the evaluation only and not responding to the immediate needs as well," Chau said.

A healthy back and forth

From Benton-Clark's perspective, she was happy that Chau pushed back and put a stake in the ground. That is just the kind of healthy back and forth she thinks the foundation needs more of, particularly so that staff become accustomed to experiencing these disagreements, not as a personal attack, but rather as working together to make different viewpoints visible to get the best product and thinking. At one point, Chau was so frustrated with Benton-Clark that Chau told her she wanted to run her over with her car—seven times. Benton-Clark said she laughed and added, "yeah, you should and you still love me."

While the two had a candid relationship, that did not mean that they always understood one another. Benton-Clark said she had broader concerns about the approach that Chau and the evaluator were taking.

"What was happening with the evaluation was that it got to the stage where it felt like there was too much emphasis on learning," Benton-Clark said. "What I mean is that there were all these meetings where it was still just about developing the learning questions. And we had talked about that we would have some tangible data, at least some qualitative data, like early stage interviews with folks just to get a sense of what was playing out on the ground. And we got to July and August and there was still no plan for that. And we had been paying a hefty sum to the evaluator for a while. I was pushing on Yen because I said we can't go to Karen and Amy [Latham] at the end of December [without any data] when they keep asking us what are some of the early stage results."

Benton-Clark also said that Chau was getting wrapped up in fleshing out an entire evaluation plan when that was not possible because the work itself was emergent and adaptive.

Chau does not see learning and evaluation as distinct as Benton-Clark does. In the summer of 2018, Chau facilitated some learning conversations with Benton-Clark and others that were aimed at surfacing their hypotheses about the locally-focused work, including what changes they expected to see. By doing so, Chau believed they could then evaluate and assess against those hypotheses, refining their thinking accordingly—the essence of learning.

"It's one of those things that the field doesn't see, the interplay between learning and evaluation," she said. "People see learning as 'knowledge acquisition,' rather than as making sense of information to inform decisions that you are trying to make. By clarifying learning questions, we can identify where the evaluation might bring data and findings to the table as an input into the learning. And then evaluation can explore whether we made the right decisions and what was the impact of our decisions, so that we can adjust our thinking and action for future decisions based on the data… in a continuous cycle of improving how we create change….I thought that Jehan and I were seeing eye to eye on this, but I don't think she fully understood what I was trying to do and then she was like 'stop this, I need you to evaluate, I need you to stop with the learning aspect.'"

Ongoing concerns about the locally-focused evaluation

Both Chau and Benton-Clark said they are still struggling with a number of issues around the evaluation of the locally-focused work. Among the issues that Benton-Clark said keep her up at night is how to keep the foundation's board engaged in this emergent and adaptive work by telling a story of what is happening in the community that weaves together a number of threads, while also making sure they are gathering the right data to inform strategy changes in real time. What are the pivots the foundation needs to make in order to further its work toward equity and what does it mean, operationally, to be community-informed? How can the evaluation help the team identify when to bring the community in to understand what *they* want to know from the evaluation? And how does the foundation define the different "communities" they are referring to, to begin with?

Chau said she worries about the ongoing work because it is so complex and there are no good evaluation examples to follow. The locally-focused work is based on an adaptive and emergent strategy that also has equity at its center and must involve the community while employing systems thinking and dynamics. An evaluation must take all of that into account and find ways to integrate learning along the way. It is a big, complicated task, particularly because the field has not developed to do this work, especially in incorporating learning into evaluation.

"There's no textbook for this," Chau said. "You're cobbling together a whole bunch of things and seeing how it works or doesn't. And using emergent learning principles or having a robust learning agenda is still very foreign to the field. It worries me that if no one else has figured this out, how am I supposed to?"

The work is sprawling and complex, happening in many different communities with program staff who operationalize their role in different and evolving ways, and with an emerging sense of what it means to work with a commitment to equity. What is reasonable to expect from an evaluation for this kind of work, and how does the foundation have to change the way it engages with an external evaluator to make sure the evaluation can keep tacking with the foundation's evolving evaluation questions and learning needs?

Equitable evaluation work begins

In 2017, one of the new cornerstones of the Colorado Health Foundation was "putting equity at the center" of all of its work. As program staff began their work with community engagement, the learning and evaluation department started helping them reflect on their equity practice. As they were doing so, Price said she and her team began thinking about what it would take for the evaluation function, too, to work a way that advances equity.

Because this was new for them, Price turned to others for help including the Equitable Evaluation Initiative (EEI), a five-year initiative to help foundations conceptualize and use evaluation to advance equity. She invited Jara Dean-Coffey, founder of EEI and the Luminare Group, which specializes in diversity, equity, and inclusion, as well as strategy development and evaluation, to run some design sessions for evaluators in Colorado, in which the foundation's learning and evaluation staff could participate. At the same time, all foundation staff are participating in broader diversity, equity, and inclusion work, which has created an organization-wide openness for sharing around equity, and exploring equitable evaluation.

As a result of that work, the team immediately began making changes in their evaluation practice, Price said.

"Every evaluation conversation we have, every evaluation deliverable we look at, our minds are churning. What do I see in here that is reflective of equity, that is getting in the way of equity?" Price said. "I actually got a question from another foundation the other day where they said, 'how do you figure out if someone knows how to do equitable evaluation?' And I said, that's not the point. We're not trying to weed out people who don't know. We're trying to weed in people who are willing to go on the journey. We're going to learn it together. Our only ask is that they're willing to do it in a very intentional way."

McNeil-Miller, meanwhile, sees the equitable evaluation approach as a "game-changer" for the foundation. She said that one of the learning and evaluation team's priorities for 2019 is to educate grantees across the state on this approach to evaluation, which she believes will free them from feeling like they do not have the staff to do evaluation and measurement. McNeil-Miller said that if you broaden the spectrum of what is considered valid, reliable, and who is considered an expert, it opens up more natural ways of doing evaluation and measurement.

But Price said the evaluation team is wrestling with a number of questions about what it means to do equitable evaluation well.

Among them is that the evaluation department started from a place where they said the *foundation* is the main user of evaluation and is doing so in order to strengthen its ability to be strategic. Equitable evaluation challenges organizations to think beyond serving itself to a broader vision of how it is also serving and benefitting the people it is interacting with in the community. Adopting that mindset brings up a number of questions that the evaluation team is still wrestling with, including how to share power with grantees over the evaluation design, methods, questions, and resources, Price said. This last question in particular is a thorny one and one that McNeil-Miller has raised.

"The other piece that I worry about all the time is equitable evaluation being treated like a checklist," Price said. "As though I just do these things and my evaluation will be equitable. And we saw a fair bit of that when we did the equitable evaluation design labs last year with other evaluators. There was a lot of 'just give me the recipe and then I can check it off and say my evaluation is equitable.' And so I worry about turning it into a checklist that evaluators tick through instead of something that they really use to critically assess their own practices and mental models."

CEO pushes back on "navel gazing" evaluations

McNeil-Miller has also told Price that she wants the foundation—and the evaluation approach—to be less navel gazing. By that, she means that evaluation questions often seem "foundation-centric" (e.g., is the foundation achieving its own goals) rather than thinking about whether the work the foundation is doing *matters* to the communities it is investing in.

"I want the impact to focus on what's the change and how are human lives better?" McNeil-Miller said. "To me it's not about a simple assessment of particular grants or a particular organization, but how do we tell a collective story of a neighborhood, a community, a state?"

Price's response is that some of those foundation-centric questions are essential in order for it to learn from its strategy and become a good strategic actor. However, she also said that when McNeil-Miller critiqued a recent set of strategy proposals, the CEO noted that the internal questions came first, and seemed lengthy, while the impact questions that she was interested came later and were far fewer. Price said that she and her team could have easily tried to "manage up" by just switching the order of questions and making it look like they were being responsive to the CEO. But she said they took on board McNeil-Miller's critique to challenge the evaluation team's thinking.

"We said, all right, what is Karen is really saying?" Price recalled. "She's really asking us if we are spending enough time and thought on what is happening in the community. It forced us to go back and reconsider the way that we were thinking about assessing impact and the role it was playing in our evaluation practice. It helped us get to a whole new thought process. What if we say, 'hey community we are here to help you change this condition?' So our evaluation practice is actually going to center on looking at the condition in the community and then look back towards the foundation and saying what role, if any, is the foundation playing in helping to improve this condition? Karen's questions led us to this idea of de-centering ourselves from the evaluation process and re-centering on the community."

Support of CEO elevates learning and evaluation

Since its inception, the learning and evaluation department has struggled with its role and authority at the foundation. McNeil-Miller's arrival has helped elevate their role, Snow and Benton-Clark said. McNeil-Miller told the department that their job is to be a mirror to the foundation and ask critical and inconvenient questions, which gives them a backing they haven't had in the past, even with supportive leadership.

"I'm protecting and supporting [learning and evaluation]," McNeil-Miller said "That means being a public advocate for them and publicly praising them. Learning and evaluation has *organizational* accountability, not just an accountability to another department. Kelci has an agenda beyond philanthropy."

Price, too, sees the role of learning and evaluation evolving. Rather than viewing themselves as an in-house consultant in support of programmatic staff and the evaluation needs they express, she views her department's role and locus of accountability differently.

"When Karen challenged us to articulate our departmental value to the organization, I was very clear that our departmental value was *not* rooted in providing support to others," Price said. "It was rooted in what it would take for us as a department to help the organization achieve the impact it wants to achieve. Our 'client' is the communities we're serving and that helps us [take a more proactive stance]."

Looking Forward

After ten years of continuous change, will that pace begin to slow down at the foundation or is it something that will continue and learning and evaluation, along with others, will have to continuously make adjustments in response?

"I think we're always going to continue to evolve," McNeil-Miller said. "Right now, that evolution we're taking is —if you know the game *Mother, May I?* —we're taking three large steps forward. In the future, we'll still be evolving but it'll be smaller steps. It'll be based on what we learned. We are unfreezing the old culture and old way of functioning. We are now moving things around in the liquid, and getting it to a gel state, and we'll eventually re-freeze it. I want it to be re-frozen as an externally facing organization that understands equity, and is strategically and behaviorally nimble and flexible and comfortable in multiple environments. Even once it's re-frozen, we'll be doing some tweaking. [But] we will stabilize, and I think we are far into stabilizing the external focus and understanding the context of the community. If people aren't comfortable in that kind of environment of constantly improving and constantly reevaluating your last decision, it will be frustrating [for them]."

Price foresees that this will be an ongoing dynamic for the evaluation team to deal with. "Once we've asked a question, people feel like it's 'done' and don't necessarily want to revisit it. Trying to get folks to constantly be circling around and checking their own thinking, and considering their actions, can be perceived as revisiting decisions we've already made and therefore frustrating to them, or us asking unnecessary questions. With Karen we have a lot more traction around the concept of adaptive strategy, but staff are still getting used to what this will actually entail in their practice.

When asked about other dilemmas she anticipates moving forward, Price said, "I'd love to believe my future is just a lovely and gentle set of rolling hills, but it will probably be more of a scream-y roller coaster ride."

Among the dilemmas she's pondering is that the locally-focused work is both an adaptive and emergent strategy, which raises a number of issues of what "the work" actually is, where to look for impact, and what is even reasonable to expect. Price is also thinking about systems in strategy, evaluation, and learning—and how to understand the set of conditions that are holding a problem in place, the variety of ways people are engaged with that system (working for or against the changes the foundation and the communities are angling for), and how the foundation's actions are helping or not helping to make progress toward the system producing different outcomes.

How to best incorporate learning remains a challenge as well.

"We are still really struggling with the idea of how you build on prior learning, instead of making the same mistakes over and over as you change strategies and staff," Price said. "This raises questions of what prior knowledge you already know and how you've implemented it, what the next question is, what it would take to answer that new question, and what knowledge management and knowledge transfer is needed. This one is really hard."

Conclusion

As Price looked back over the last several years of continuous change at the Colorado Health Foundation, did the evaluation department need to go through all of the ups and downs and twists and turns, including building trust and a strong role for itself, in order to get to the place where it is now? Or was there a simpler path for someone in her role in both reacting and responding to change?

Price is not sure, but she said that an initial struggle was getting both the foundation leadership to let go of old mental models, such as measurable results, and the evaluation field, including herself, to do the same. In the case of the evaluation field, she said that those mental models are that the field still views itself as objective data analysts. Instead, when working at foundations, external evaluators and evaluation directors need to be intimately involved in strategy, because learning should be at the core of strategy and evaluation is a key input into learning.

"I don't understand how evaluation can be separated from strategy in an organization, and yet a lot of organizational structures and thinking in foundations separate these things out and treat evaluation as though it is *programmatic* evaluation," Price said. "As evaluators in philanthropy, we need to evolve our own practice and understand what we need to know and do, and that is fundamentally different than how most of us were trained. I feel like if we could get better in philanthropy at just admitting there's a different way we need to do the work and being really diligent about creating an opportunity for people to learn into that, then it would accelerate practice."

"I'm still puzzled by foundation evaluation teams who are only there as a support resource," Price said. "Because then you're letting everybody else drive evaluation and learning and how it works in the organization. You're only getting to insert yourself when somebody asks you to come in. And then you're in this kind of client relationship where they've invited you in and you've got to be all solicitous and do what they want and you can't say, 'well that's actually really not a useful question you're asking. Let me suggest something else.' I have this feeling that part of what we need is the evaluation field or function seeing itself in a more proactive way, supporting the change process

	Starting with outputs (2009-2012)	Getting learning and evaluation in place (2013-2016)	Evolving to support adaptive strategy, and community focused approach (2017 +)
Evaluation Practice	***<u>What</u>: Movement from almost exclusive focus on output metrics with little eval, to eval studies at the core of how we assess progress. Evolution of eval approaches to align with adaptive strategy and addressing question of what difference we are making.***		
	What: Output metrics at the center - Set of standardized output metrics reflecting general strategic direction - Mining and aggregation of data intended to produce insights about strategy - 2 evaluation studies between 2008 and 2012	**What: Evaluation studies at the center** - Standardized metrics remain for grant-level - Launch evaluation studies to inform strategic decisions (5 in 2013; up to 23 by 2016) - Evaluation studies focus on initiatives - Evaluation studies link explicitly to next decision point - Not everything gets eval – prioritize key Qs - Embed eval planning into strategy devp't	**What: How to evaluate adaptive strategy** - 23 evaluation studies in 2017 - Org transitions delay launch of new studies and evolution to cross-cutting eval studies - Prioritize strategy-level retrospectives covering 2009-2017 to inform upcoming transitions - Start integrating new CEO's question of "What difference are we making?" into all evaluation studies
Learning Practice	***<u>What</u>: Want to be a learning org, but initially no structures for org learning – only personal learning happening. Adoption of learning practice to support better conversations. Integration of learning and evaluation moments aligned with decision-points into planning.***		
	What: Personal learning - Personal learning dominates - Lack of explicit practices to create learning across people or teams - Sharing of info centers around yearly presentation of output metrics to Board	**What: Learning within and across** - Adoption of Emergent Learning approach - First attempts to learn within specific bodies of work and about cross cutting Qs - Learning moments part of planning discussion; intentional alignment with decision points and evaluation reporting - Test strategy-level learning discussions	**What: Learning for adaptive strategy** - Major org changes shift learning back to "as needed, where needed" - First impact reports on initiatives that are ending, including judgment of 'success' - Retrospective studies provide important moments of reflection for the teams
Org Context That Impacts Evaluation and Learning	***<u>What</u>: Org values learning and evaluation but lack practices to make it happen. Responsive grantmaking and lack of clarity about what change is sought makes it hard to go 'beyond the grant'. Several major changes in executive leadership and strategic approach over 5 yrs.***		
	What: Responsive strategy - Earliest strategic plans ('06) note evaluation and learning key to success - Responsive grantmaking coupled with lack of clarity about change sought makes it difficult to do evaluation 'beyond the grant' - Size and scope of foundation makes comprehensive grant-level evaluation impractical (active grants each year = ~500)	**What: Moving to initiatives** - Major strategic refresh in 2013 provides opportunity to shift learning and eval approach - Organizational move towards more structured grantmaking (e.g., initiatives) creates different evaluation opportunities - CEO departure late 2014; new CEO Sept 2015. Transition state = no major changes	**What: Shift to community engagement and equity** - Major shifts to program staff roles in how they engage with community - Major shifts in approach to grantmaking - move away from initiatives - New strategic direction developed during 2017 for 2018 implementation - Org shifts towards focus on equity in practice and outcomes

	Starting with outputs (2009-2012)	Getting learning and evaluation in place (2013-2016)	Evolving to support adaptive strategy, and community focused approach (2017 +)
The: "So what?"	***So what*: Lots of support for learning and evaluation with a fairly blank slate on which to build. Learning practice and eval studies quickly demonstrate value, increasing buy-in. Org shifts create opportunities to implement new learning and eval practices, but multiple shifts also require continuous evolution of the model and keep it in the 'development' vs. 'maturity' phase.**		
	So what: Little learning, little evaluation - Lots of interest in evaluation and learning - Largely blank slate on which to build new model - Staff realize they are unable to answer questions about strategy, what's working, and lack data to inform real-time practice - Recognize need to build our muscle around learning - Recognize need to get data on what is happening with the work and what it's accomplishing - Willingness to fund evaluation studies	**So what: Growing learning and eval practice** - Org strategic changes provide opportunity to embed planning for eval and learning into development phase - Focus of evaluation stays on initiatives because that's major unit of work - Buy-in for evaluation studies remains strong as they show early value - Learning improves markedly, but still sporadic depending on readiness and interest - Emerging recognition that lack of documentation about strategy and activities impacts learning across time	**So what: Some conclusions…and back into refresh** - New program staff roles and place-based work offer unanticipated opportunities for embedding learning practices into staff training and development processes - Ability to report on closing initiatives resonates strongly with staff and Board, seeding interest in assessing success and discussing findings - Major org changes cause transition state, slows plans to evolve learning and eval work - New strategic direction almost in place, renewing org interest in doing more regular adaptive learning about strategy
Now what	***Now what*: Start with getting away from output metrics, put a learning practice into place, and get eval studies started. Integrate planning for learning and eval into org strategic planning, and align with key decision points. Start evolving for adaptive strategy.**		
	Now what: Put learning and eval at the center - Evolve the evaluation model to put evaluation at the center, not output metrics - Develop an approach to learning intentionally - Start assessing our work through evaluation	**Now what: Revolving around initiatives** - Position evaluation at the strategy level, not initiative level - Integrate learning over time, not just one-off learning sessions - Improve knowledge management to support learning over time	**Now what: Evolving to adaptive strategy** - Evolve our evaluation and learning approaches to better match the type of strategy – adaptive, emergent, etc. - Bring learning "back into the center" and institutionalize more broadly - Redevelop the evaluation model to accommodate new structure of the work (e.g., not all initiatives anymore)

APPENDIX A

	Starting with outputs (2009-2012)	Getting learning and evaluation in place (2013-2016)	Evolving to support adaptive strategy, and community focused approach (2017 +)
Activities	***Start small and organic, knowing what we want to accomplish (see principles). Test out eval and learning activities, iterating across time. Use learning practice on ourselves to improve. Didn't lock in structures and practices too early – be nimble, experiment.***		
	Evaluation Activities: - Analysis of output metrics - Quality checks on grant metrics - Two evaluation studies **Learning Activities:** - Yearly report to Board on outputs - Test some summaries of progress for key initiatives	**Evaluation Activities:** - Launch theories of change - Launch Measurement and Learning planning tool - Launch evaluation studies broadly **Learning Activities:** - Adopt Emergent Learning framework - Facilitate learning conversations - Test 'learning pauses' embedded in initiatives - Test grant and org level learning agendas - Test strategic debriefs - Test documentation of strategic activities	**Evaluation Activities:** - Theories of change - Measurement and Learning planning tool - Evaluation studies - Retrospectives of strategy (2009-2017) **Learning Activities:** - Test 'closeout' initiative reports - Revised templates to track strategic thinking, activities, and outcomes of each strategy - Test infographic summaries - Using learning activities/frames to support program officer role transition
Team structure	***Early growth, then team structure stable over time. Department name change in 2017 to indicate importance of Learning and Evaluation (with learning deliberately in the lead, and evaluation as one tool to do that).***		
	1 staff (2008) to 4 staff (2012) Separate from program department Name: Research and Evaluation	4 staff Separate from program department Name: Research and Evaluation	4 staff Separate from program department Name: Learning and Evaluation
Team scope	***Team scope increases markedly over time, moving from focus on data-analysis to designing evaluation studies and learning opportunities. Requires new skills and understandings of how to adapt learning and eval to strategy-level assessments, including situations of complexity, systems, adaptive and emergent strategy, developmental evaluation, etc.***		
	• Checking standardized output metrics for every grant • Summarizing data from output metrics • Overseeing 1-2 evaluation studies • Some direct TA to grantees around evaluation and measurement • Monitoring state surveillance data	• Supporting program staff around strategy development • Theory of change development • Scoping and implementing evaluation studies with external evaluation partners • Facilitating learning sessions • Prioritizing focus of learning and evaluation	• Supporting program staff around strategy development • Theory of change development • Scoping and implementing evaluation studies with external evaluation partners • Facilitating learning sessions • Prioritizing focus of learning and evaluation • Analyzing strategy and mapping on appropriate evaluation strategies (e.g., adaptive, emergent) • Knowledge management (stewardship role)

THEME 1: Evolving views of strategy

Over the course of the case, the view of the foundation's board and leadership evolves considerably with respect to what constitutes "good strategy" and what it means to be strategic.

- How would you characterize what the foundation believes "strategy" is in each phase of its evolution? What counts as a "strategic decision"?
- For each successive re-imagining of strategy, what are the underlying assumptions about how social change happens?
- What are the implications of this evolving view of social change for the roles and responsibilities of program staff? What does it imply about how the foundation views the role of grantees in social change (and in relation to the foundation)?
- What prompts the foundation board and leadership to change its view of the foundation's role in social change? What do they hope each new iteration of strategy will make possible that previous ones did not?
- What can we infer about what it means to be a "high performance" foundation in each phase? What are staff accountable for? What about grantees? How does this change as the foundation decides to focus on equity?
- How does <u>your</u> foundation conceive of "good strategy" and what does it imply about your foundation's underlying mental models of social change?

THEME 2: Evolving role of evaluation and learning

As the foundation's ideas about strategy evolve and new staff come into evaluation leadership roles, we see similar evolution in—and tensions between—views about evaluation and learning.

- How does the changing nature of "strategy" affect the focus of inquiry for the evaluation unit? In other words, as the foundation re-imagines its role in social change, what new questions and tasks does it raise for the evaluation unit? How does it change what they're busy with on a day-to-day basis?
- What does it mean for the evaluation team to treat strategy as the unit of analysis?
- How do the roles and authority of evaluation staff change over time? What tensions does this create vis-à-vis the roles and authority of program staff? How do staff and leadership navigate those tensions both successfully and unsuccessfully?
- In the later sections of the case, what does the evaluation unit believe its role in strategy development—and ongoing strategy—is? How does this differ from other views of the relationship between strategy and evaluation?
- As the evaluation team adopts a more explicit focus on learning, what can we infer about what they believe 'learning' requires? What are alternative ways of conceiving of a learning role that might offer different kinds of value to the foundation's work?
- We learn toward the end of the case that the evaluation unit is beginning to see itself as accountable to the communities with whom the foundation and its grantees are engaged. What does this mean and how is it different from their implied line of accountability in the past? How might it change their focus, their priorities, and even their day-to-day work?
- At your foundation, how do program and evaluation and/or learning staff negotiate their roles and scope of authority? What tradeoffs and tensions do they navigate?

THEME 3: Navigating organizational change

The foundation undergoes a rapid increase in size, a significant transition in leadership, two large-scale strategic re-orientations and corresponding redesign of program staff roles and competencies, and an overall departmental restructuring. In addition, its newest CEO brings an explicit commitment to equity.

- How does each phase and type of organizational transition affect staff disposition and behaviors in general?
- More specifically, how does it affect their mindsets, bandwidth, and willingness to engage in different types of evaluative thinking and work? What does this suggest about how evaluation leaders should engage with staff during periods of upheaval?
- The evaluation team decides to play a proactive role in the change process by assisting with the development of new strategies. What are their assumptions about the evaluation team's "value-add" and responsibilities in the development of new strategies? How do others in the organization view the team's value and responsibilities in this process?
- At one point in the case, the evaluation team decides to promote the use of theories of change (TOCs) as the tool for articulating new strategies during the transition. What are they hoping to accomplish with this and how does it fit with the phase of transition the foundation is in? What tensions play out around the TOC process and what might be driving these tensions?
- What do you believe is the appropriate role of evaluation staff during organizational transition and strategy re-design? What should they be responsible for?

THEME 4: Managing Complexity

Foundation staff experience a high degree of uncertainty as they go through several organizational transitions and ultimately embark on an approach to change that is intended to acknowledge the complex and emergent nature of the problems they hope to address.

- Where in the case can we see staff struggle with the tension between the desire for certainty and control versus the "unknown" or uncontrollable nature of complex change?
- What other tensions does systems change require us to navigate?
- What reactions to uncertainty —or strategies for managing it—do we see playing out at the individual or organizational levels? What are their consequences? What alternative responses might the foundation have to uncertain conditions?
- How does the evaluation team think about the role of evidence and data in the context of complex change work?

THEME 5: Foundation/Evaluation Consultant Interaction

We see examples in the case of evaluation consultants working to provide evaluation work that the foundation finds useful and adds value to its work. One highlighted evaluation discovers mid-stream that an initiative is not unfolding as intended and is asked to revise the evaluation approach so that it might produce insights that will be useful to the foundation in its new strategy, which is still under development.

- Given that the evaluation of the HEAL initiative is hitting the ground right as a new CEO has arrived and charting out a new direction, what interest and attention do we suspect foundation staff have for the "old" work the evaluation team is evaluating?

- How do staff and the evaluation team approach the redesign of the evaluation? How would you characterize why this redesign didn't work as expected? In hindsight, what might the evaluation team have done, if anything, that would have been more valuable to the foundation?
- What would the evaluation team have to understand about the foundation's new strategy to be able to effectively adapt the evaluation mid-stream to be useful going forward? Why might the foundation be hesitant to communicate the relevant information to the evaluators?
- What conditions or norms of interaction need to be in place in the relationship between evaluation consultants and foundations to ensure that work stays relevant even amidst transition? How do foundations and evaluators need to interact differently to make sense of data and inform strategy in powerful—or even transformative—ways?
- What is the nature of the relationship your foundation has with external evaluators? Is there a gap between what you want from these engagements and what you tend to get? What are the institutional or cultural habits and routines that aid or hinder consultants from effectively adapting to your evolving thinking and emerging questions?

This teaching case was produced as a part of the April 2019 Evaluation Roundtable Convening held at the Margaret A. Cargill Philanthropies in Eden Prairie, Minnesota. We are deeply grateful to the Colorado Health Foundation and all those featured in the teaching case for their candor. They gave philanthropy a gift in offering their experience up to a broader audience to reflect on.

The Evaluation Roundtable is a project of the Center for Evaluation Innovation. It is a network of foundation leaders that aims to improve evaluation and learning practice in philanthropy. Founded in 1989, the network includes leaders from over 100 foundations in the U.S. and Canada. It is a preeminent resource for information and ideas on how foundations approach evaluation and learning.

The network is based on the idea that helping people connect deeply with new ideas and with each other will speed the development and spread of solutions to evaluation challenges in philanthropy. Through joint problem solving and knowledge creation, participants join forces in making sense of and addressing dilemmas they face individually or collectively.

The teaching case model is a cornerstone of the Evaluation Roundtable Convening. If you are foundation evaluation staff and your foundation is interested in being featured as a future teaching case, please reach out to Tanya Beer and Julia Coffman, co-directors of the Evaluation Roundtable. They can be reached at tbeer@evaluationinnovation.org and jcoffman@evaluationinnovation.org.