EDUCATION CONNECTIONS NEWSLETTER

Welcome to ICF *Education Connections*

Each year government agencies spend millions on technical assistance designed to improve education, health, and other social services. But how much do we know about the impact of these dollars? Evaluation of the interventions these agencies conduct to improve services could begin to answer this and related questions. In this issue of ICF *Education Connections*, we consider one complex effort to implement capacity building and evaluate its effectiveness, and hear from one of ICF's experts on the topic.

- Launched in 2014 by ICF, the Capacity Building Center for States, a service of the Children's Bureau, embedded evaluation into management and service delivery from the beginning.
- Caitlin Howley, ICF Fellow, reflects on the topic of evaluating technical assistance. Dr. Howley leads education research and program evaluation studies and provides capacity building technical assistance to educators.

From early education to K-12 and postsecondary schooling, ICF gives clients the insights needed to turn research into practice. Visit our website for more information about ICF's work in education. Want to share *ICF Education Connections*? Click here to subscribe.

Empowering Agency Improvement through Evaluation of Capacity Building

From day one, the Capacity Building Center for States—a service of the Children's Bureau—intended its evaluation to assess whether its services are delivered as designed, whether services are achieving outcomes, and to inform efforts to improve. The Center for States provides coaching, consultation, facilitation, publications, and trainings to help state and territorial public welfare agencies build their capacity to implement effective child

welfare practices in order to improve safety, permanency, and well-being for children, youth, and families.

A new federally funded capacity building center launched by ICF in 2014, the Center for States partnered with evaluators, practitioners, and experts throughout the development of its Capacity Building Practice Model, which guides its support to child welfare agencies. The Center designed its model to create consistent practice by setting up an explicit, shared set of behavioral expectations for service delivery teams. The model describes these behaviors in concrete terms and categorizes them into three levels of desired, developmental, and insufficient practice. The model creates a roadmap for evaluators as they develop methodologies for determining the level of implementation fidelity. In short, the model provides an opportunity to understand the extent to which capacity building, as a fully developed intervention, can be operationalized and implemented as well as evaluated to understand its effectiveness.

By embedding evaluation into management and service delivery from the beginning, the Center for States also created the opportunity for a nimble evaluation approach that could evolve with the Center. The Center capitalizes on data visualization software and a commitment to continuous quality improvement to regularly use feedback to adjust and innovate service delivery. The Center also captures evidence of changes in organizational capacity through evaluation capacity building, conducted by child welfare evaluation experts across the country. Each of these experts serves as a coach to help states collect data on progress, implementation, and goal achievement. Specifically coaches:

1. Provide capacity building services that
    - Facilitate outcomes-focused thinking (e.g., logic models)
    - Develop an evaluation plan with the state
    - Build evaluation, data, and analytic skills to monitor progress
    - Ensure evaluation approaches align with state evaluation capacity
    - Assist the state in using data to reflect on progress and adjust

2. Facilitate evaluation data collection that
    - Facilitates methodology that leads to high quality data
    - Supports the state in collecting data to show achievement of short-term and interim outcomes
    - Supports the state in understanding and analyzing data

The Center for States evaluation strives to use innovative evaluation methods to address a unique set of challenges driven by high service delivery volume, a widely diverse set of capacity building services and contexts, and the use of a multitude of data sources and types. For further information please contact Christine Leicht, Center for States Evaluation Lead (Christine.Leicht@icf.com) or Brian Deakins, Children's Bureau Contracting Officer (Brian.Deakins@ACF.hhs.gov).

ICF Interviews Caitlin Howley on Evaluating Technical Assistance

Recently we interviewed Caitlin Howley about evaluating technical assistance. Here are excerpts from that interview with Dr. Howley, director of the Appalachia Regional Comprehensive Center and ICF Fellow.

Interviewer: Today, we are speaking with Caitlin Howley, director of the Appalachia Regional Comprehensive Center and our topic is evaluation of technical assistance. Why is it worthwhile for us to evaluate the technical assistance that we provide to our clients?

Caitlin Howley: I think that's a great question to ask because a lot of times folks feel as though their budgets are already constrained and why would they devote additional resources to evaluation. So I think it is a really pertinent question generally, but in particular for clients as well who are pondering whether or not to engage an evaluator.

I think there are three major reasons for conducting evaluation of technical assistance. The first has to do with improvement—to learn better how to do what it is that we do. So development and formative evaluation really help us understand how we could improve our services and the processes and tools that we use to assist clients.

The second big reason for evaluating technical assistance is to document/investigate whether our assistance is actually helping clients—asking questions about whether we're helping them achieve what it is they want to achieve, whether they're gaining the knowledge/skills/capacities that they require to do their work or achieve their goals.

And a third reason to conduct evaluation of technical assistance is to contribute to what is really a very scant empirical literature on technical assistance. The federal government depends on technical assistance as a key venue for supporting school reform and school improvement and a lot of different kinds of support services fall under the aegis of technical assistance, but there is very little careful documentation of what goes on during technical assistance and relatively little guidance, therefore, about how to move through the process of providing services that really help clients achieve what they need to achieve.

Interviewer: The second purpose you mentioned—documenting, investigating, helping clients achieve what they want—one of the challenges is that clients are going to be after all kinds of different goals for what they hope to achieve through technical assistance?

Caitlin Howley: That's absolutely true and one of the arts that technical assistance providers bring to each engagement with a client is their ability to help clients focus in on something that is clearer, more achievable, more closely tied to their overall goals. You know, in some ways beginning a project, a technical assistance project, is the most important phase of that technical assistance project because that's where the boundaries are defined, that's where the scope is defined, that's where the shared values and norms are defined for the work that we do together with clients. It's a critical piece of the process

and that's where rapport gets built as well, so that you can talk honestly and critically about the issues at hand.

Interviewer: So that early phase becomes a place where you would start to think about the ways you'd go about doing the evaluation. What kinds of evidence are you going to be looking for when you're doing this kind of evaluation?

Caitlin Howley: The answer isn't particularly satisfying but it is apt, which is that it depends. It depends on the scope of the technical assistance project, on the purpose of the evaluation, and you know, even just on those two dimensions there is lots of variation. There are a couple of models or approaches out there so, for example, the Donald Kirkpatrick model has four levels of evidence that one might look for in an evaluation.

The first level is *reaction*. Asking questions about the extent to which participants find the technical assistance engaging, relevant, [and] useful. And there you might ask for self-reports, ratings of how satisfied they were with the training, how relevant it was to their work situation. And that's a very immediate and basic level of evidence.

The second level in the Kirkpatrick model is *learning*, so these are evaluation questions about the degree to which the participants gain the intended knowledge, skills, attitudes, confidence, or commitment based on their participation in the training. And again, here you can use self-report, but we can also use knowledge tests, which are essentially more psychometrically sound measures of what people actually know, not just what they say that they know. Other kinds of evidence might come from observation. Can people demonstrate their knowledge or application of a skill? Are there artifacts that we could review, for example?

So, as you move through this Kirkpatrick model series of levels, you can see that not only is the content becoming more complicated, the data collection method or instruments or tools become more complicated as well, so by the time you get to Level 3, which is *behavior*, these would be questions about the extent to which participants apply what they've learned

during technical assistance when they're back on the job, for example. Again, self-report is a standard way of getting that information, but document review, observation, reports from others as well. For example, in doing 360-degree evaluation, you can gain a picture from reports from people who are colleagues of the people in question, who supervises them, or whom they supervise.

The fourth level is *results*, and this is more about the degree to which targeted outcomes happen as a result of the technical support and assistance provided. Observation, document review, evidence of how people's behavior changed policy/practices/programs, whether those policies were in fact developed and then implemented, and too, you can look here at how those policies/practices/programs that were implemented then affected larger groups of people, and that's the sort of result you characterize as impact, for example. And there you might be looking at program data or census data. With those sorts of impact level data, you can also do more complicated designs like randomized control trials or quasi-experimental designs. It really depends on the project though, if you're able to get to that point. Not all projects warrant that level of investigation. So the Kirkpatrick model is one very clear, straightforward way to approach collecting different kinds of evidence.

The second approach would be a case study approach, which is much more narrative and is about documenting the story of what happens during technical assistance and what happened as a result of it. And here it's a longitudinal sort of design that employs multiple methods and perspectives to understand a bounded case as comprehensively as possible.

A third approach is more process-oriented. Here you might rely a lot on project data, or observation, or engaging participants in journaling, reflecting on their own learning and experiences as they progress through the technical assistance effort, and that might be a way to answer questions that are more developmental or formative rather than summative. So it's more about what is happening during technical assistance and how are participants making sense of it, versus the Kirkpatrick or case study model which may focus more on

summative aspects of a technical assistance project, meaning what happened as a result of the work performed.

Interviewer: You've mentioned before that the literature on this topic is pretty scant. I'm guessing that there aren't too many examples that take this, say, to the fourth level of Kirkpatrick's structure.

Caitlin Howley: There are very few. I only know of a handful of randomized control trials that investigate a technical assistance project. And those tend to be a distinct kind of technical assistance project. Those tend to be technical assistance projects that are focused on helping clients implement a specific evidence-based practice or program. In many ways that makes the control trial approach better accomplished. If the project is more diffuse or expansive or adaptive, then it is harder to specific what "it" is under investigation and to do a very rigorous kind of experiment.

Interviewer: Are there other key factors you'd want to point to that people need to consider in designing an evaluation of technical assistance?

Caitlin Howley: Yes, there are a couple of things I think are important to keep in mind during design of evaluation of technical assistance. I think, first, it is important to recognize that it is really unlikely that you'll be able to attribute a causal relationship between your technical assistance services and any eventual outcome. Instead, I would suggest that it is more appropriate to think of technical assistance as contributing to the outcomes in question. So that's a first thing I think it's important to contend with.

Secondly, I would recommend that unless you're evaluating a technical assistance project that is helping clients implement something that is well proven already in the field, an evidence-based program for example, it's important to acknowledge that a lot of technical assistance work is adaptive. It's not just technical. It's adaptive in the sense that the problems clients face are often pretty complex. There are no clear right or wrong answers or approaches to those complicated problems and lots of course correction might be required

as you work and learn together with your client. So defining up front what the technical assistance will be can be challenging and frustrating for an evaluator who is perhaps more accustomed to evaluating prescriptive programs. Technical assistance often is not that.

Another important thing has to do with data. You know, technical assistance efforts can often be sort of expansive or even ungainly, and it's important to get comfortable with the idea that you can't evaluate everything. You can't collect all the possible data out there about a technical assistance effort. You can't be an evaluator in all kinds of places at once. We haven't figured out yet how to break the space/time continuum for that.

So you have to be okay with making some choices. And relatedly, you have to be okay with balancing the need to collect data with the need to limit data collection burden for participants. I think that's probably the case in any evaluation project but in technical assistance I think it's important to keep that in mind as well, particularly since things evolve and you may be doing needs assessment with your clients as well. So that's just something to keep in mind.

And then I guess another important thing would be to build the capacity of your technical assistance providers to regularly collect and use informal means to gauge how things are going for the client. So there are tools like a simple plus Delta activity at the close of each meeting or engagement. It's a great way to just check in. What was good about this? What could have been better? And that way you can quickly get some formative evaluation data. It's embedded in your meeting. It's not awkward. It's not just tagged on later after the event, and it can be used meaningfully to course correct if you need to.

Interviewer: It's easy to understand as you're talking why there might be a scant empirical literature on technical assistance. This sounds like a very complex kind of evaluation.

Caitlin Howley: It is. You know, technical assistance itself is more art than science, and I suspect it always will be, because it is a social and human enterprise and the kinds of

questions that people are encountering in the context of these organizations, which are themselves nested in community, and policy, and socio-economic contexts that are often shifting and constrain and also enable people's possibilities, but you know, ultimately it is the job of the technical assistance provider to help clients make some sense of that and sense-making is, again, a social process—more art than science.

And you and I just read a book recently called *Humble Consulting* (Edgar Schein, 2016), and I appreciate the main message it conveys, which is that technical assistance providers have a certain set of skills but one thing that they don't have is a repository of clean answers to every problem, and it's important as a technical assistance provider to come into an organization or agency or group of clients humbly and with a clear understanding that they are the experts on their contexts and agencies, not you. You're there to help them find their own solutions.

***Interviewer*:** Are there examples of approaches to evaluation of technical assistance that you'd want to point to?

Caitlin Howley: The center I direct, the Appalachia Regional Comprehensive Center [ARCC], its mission is to provide technical assistance that builds the capacity of state education agencies to undertake a wide array of programming—school improvement, and support for teachers, and college- and career-readiness standards, etc. And one of the things that is happening in the ARCC in terms of evaluation is that we have traditionally developed some measures of specific state department of education capacities, whether they were human capacities (like knowledge and skill), organizational and structural (having to do with relationships among units and also policies and procedures, those sorts of things that exist regardless of the specific people who are in place to implement them), and we're now thinking more about political capacity (the ability of state departments of education to form coalitions with stakeholders, to advocate for particular policies or programs that they think are important to education and their states). That's traditionally what we've done via data collection methods like surveys and phone interviews. And we're moving now more towards also documenting policy, program, and practice outcomes that result from client

actions that were in turn supported by ARCC technical assistance. So it's more about following the through-line. Here is what the state department of education said it needed to achieve, here is how the ARCC helped, then the state department of education did X with a result of Y. So in some ways it's a mash-up of measuring capacities as best as we can and then telling the story in a way that a case study might. And in fact, we for several years did case studies. There were some challenges there partly because it is difficult to tell a story that takes place over five years. It's also difficult to tell a story when the technical assistance must shift as client priorities change or new leadership emerges, etc. But I think there is a lot of value in an evaluation of technical assistance, telling the story of that through-line from technical assistance engagements all the way through to more diffuse impacts in the department and out in the state.

Interviewer: What you're saying does help me see your choices about what data to collect are critical in terms of trying to produce an evaluation that actually does provide you and the client with some worthwhile understanding of what has gone on.

Caitlin Howley: Yes, and it's also important then to, early on, find in collaboration with the client what the goals are. What is the purpose of this technical assistance project and what would it be most meaningful to try to document or measure during this evaluation?

***Interviewer*:** Is there anything we haven't talked about around this topic that you would like to say anything about at this point?

Caitlin Howley: I think there is lots of room to learn here and that's why to my mind, you know, that important and maybe under-championed reason for doing technical assistance evaluation is to contribute, to begin to build a literature that is empirical, not just prescriptive or descriptive, about technical assistance and the complexities of it, the nuances of it. I think we even have some difficulty describing the theoretical assumptions between various technical assistance approaches. It seems to me that might be an

important place to start. So when we're beginning to engage with a client for example, what assumptions are we making about how their organization is likely to function?

Interviewer: Thank you so much for your time today, Dr. Howley.

Review Past Issues

Want to share Education Connections? Click here to subscribe.
